

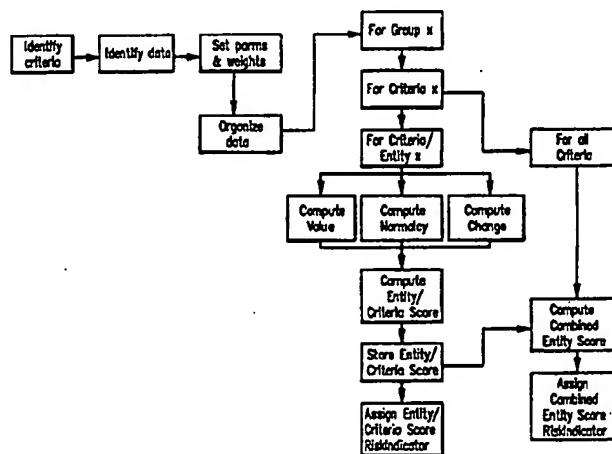


AE

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G06F 15/00, 17/60, G06G 7/52		A1	(11) International Publication Number: WO 97/00483 (43) International Publication Date: 3 January 1997 (03.01.97)
(21) International Application Number: PCT/US96/10352			(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(22) International Filing Date: 14 June 1996 (14.06.96)			
(30) Priority Data: 08/490,984 15 June 1995 (15.06.95) US			
(71) Applicant: FRAUDetect, L.L.C. [US/US]; Suite 312, 10400 Eaton Place, Fairfax, VA 22030 (US).			
(72) Inventor: COFOD, Robert, K.; 282 Prince's Lane, Harwood, MD 20776 (US).			Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.
(74) Agent: GROSSMAN, Jon, D.; 2101 L Street, N.W., Washington, DC 20037 (US).			

(54) Title: PROCESS AND APPARATUS FOR DETECTING FRAUD



Not a parts system
Driver parameters/ values/ are
Programmatic/ function/ priori/

(57) Abstract

A system and method for processing transaction data in filtering and targeting potentially fraudulent entities within a particular industry relies upon client data input apparatus for inputting client/provider/entity data records containing data elements for measuring an entity's performance according to pre-selected criteria, where entity data, criteria data and time data are organized and stored in three-dimensional array. A pre-processor (110) generates a score for each entity-criterion (1..n) by generating a value statistic (v), a normalcy statistic (n), and a change statistic (d), weighting each statistical value, and summing the values. An entity scoring generator (30) generates a single entity score for each entity by weighting each entity-criterion score and summing the entity-criterion scores. The entity scores are compared to pre-defined threshold values for detecting those entities most likely to be engaging in fraudulent activity and a graphical display entities according to level of risk.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

PROCESS AND APPARATUS FOR DETECTING FRAUD

5

FIELD OF THE INVENTION

10

The invention relates to a process and apparatus for analyzing client data against pre-selected criteria over a pre-determined period of time in order to identify those individuals or entities engaging in activity indicative of potential fraudulent behavior within a particular industry.

15

BACKGROUND OF THE ART

20

25

25

Many industries, such as medical and banking, suffer enormous losses due to the theft of money or services through fraudulent means, such as submitting insurance claims for unseen patients or forging checks. In the medical industry, for example, it is estimated that approximately \$100 billion of U.S. annual health care expenditures is stolen through fraud or abuse. Even more alarming is that some experts believe this figure only represents the portion of which we are aware. The actual figure may be significantly higher, perhaps as much as twice the amount.

30

The problems surrounding fraud detection are many. Since most companies maintain relatively few investigators, limited resources are available to perform the analysis and decision-making necessary to accurately attack the problem and ensure measurable results. In addition, the possibility of wrongly accusing a client of fraudulent behavior adds an extra burden of caution and proof. Confidentiality, government involvement, market

competition, and judicial processes add other limits to the options for attacking the problem.

Other problems in fraud detection include highly varying types of targets, possible scenarios, and innovative fraud techniques requiring heavy reliance on skilled human analysts and decision makers to detect, recognize, define and initiate actions. In addition, each industry, such as the aforementioned medical and banking industries, generates a massive and dynamic quantity of information from which relevant "target" information must be extracted. Furthermore, there is a time sensitivity factor in processing and acting upon newly detected targets or instances of fraud.

15

Due to the above-cited difficulties in detecting those persons engaging in fraudulent activities, billions of dollars worth of stolen money or services have largely been ignored or passed on as a cost of operation.

20

Most cases of fraud are discovered as a result of tip-offs or indications from industry program integrity reviews. These discoveries have often been cited as being only the tip of the iceberg, yet most investigative staffs are already inundated with the cases on hand.

30

The problem is not unlike another well-known problem in military intelligence and targeting circles. It is known as a "target rich environment," where the number of possible targets far exceeds the resources available for attack. The military problem, like the one faced in combating health care and banking fraud and abuse, is fundamentally one of resource management. The problem

requires a method or system by which an industry may increase ensurance that the resources available to it are being spent on targets with the highest potential payback. Consequently, there is a need for processing, ordering and analyzing the volumes of transactional data, such as 5 billing information and insurance claims, associated with a particular entity, such as a physician working within the health care industry, and extrapolating from such information whether the client is engaging in fraudulent behavior. This need for specifically targeting 10 perpetrators of fraud or abuse becomes more readily apparent in light of the fact that there may be tens or hundreds of thousands of clients per industry. Accordingly, the present invention answers that specific 15 need.

SUMMARY OF THE INVENTION

It is therefore apparent from the above that there exists a significant need for a system and process which 20 effectively uses transactional data in a manner that identifies potentially fraudulent behavior and areas of greatest need for investigation. It is therefore one of the objectives of this invention to provide a fraud pre-processor detection system. Fraud can be detected for 25 many industries dealing with a large number of clients, such as the aforementioned medical and banking industries.

Another object of this invention is to provide a process for analyzing large volumes of transaction data for 30 the purpose of detecting indications of fraud.

Another object of this invention is to provide a system and process which automatically prioritizes those

clients or entities which are most likely to be engaging in fraudulent behavior.

5 Another object of this invention is to provide a system and process for establishing thresholds indicating fraud.

10 An additional object of this invention is to provide a system and process for filtering, processing and displaying data in a manner where it may be used as an analysis tool to support the human decision making process.

15 Another object of this invention is to provide a system and process for organizing data in a three dimensional array consisting of pre-selected criteria along the x-axis, groups of entities along the y-axis, and time units along the z-axis, thereby allowing entity-criterion analysis, that is analyzing each entity within a group (a single chiropractor out of a group of chiropractors) 20 according to each criterion (e.g., the number of patients seen by the chiropractor, or the amount of claims submitted to the insurance carrier), over a pre-selected period of time units (e.g., days, months or years).

25 Yet a further object of this invention is to provide a system and process for generating statistical values for each entity-criterion to measure Value (v), or the relative importance of the entity with respect to the entire group, Normalcy (n), or how the entity conforms to 30 the normal behavior of the group for a particular criterion, and Change (d), or the amount of change in activity by the entity over a period of time.

It is also an objective of this invention to weight the statistical values for Value (v), Normalcy (n) and Change (d) according to the relative importance of each value within a criterion with respect to targeting entities 5 engaging in fraudulent activity.

Yet a further object of this invention is to add all the statistical values for each entity-criterion together to create a single entity-criterion score, weight 10 the score according to its fraud targeting potential, sum all the individual entity-criterion scores to form a final entity score, and compare that score to a threshold value to determine the risk of fraudulent behavior.

15 A further object of this invention is to display all the entities under analysis in a manner that allows easy identification of those entities engaging in behavior most likely indicative of fraud.

20 Another object of the invention is to provide a list of actions to take against those entities suspected of fraudulent behavior so that a human analyst may select the appropriate action.

25 It is a further object of this invention to provide a system and process for monitoring those entities suspected of engaging in fraud once an action has been taken against the entity, and determining whether the behavior of the entity changes, and measuring such change, 30 thereby providing some evidence that fraudulent behavior was accurately identified and corrected while also providing a mechanism by which the weighting of the relative importance of various criteria, entity-criterion

scores and entity scores can be modified to more accurately detect the presence of fraudulent behavior.

5 It is yet a further object of this invention to provide a system and process for generating a change meter which can be translated into dollar amounts to calculate savings to a particular industry.

10 Briefly described, these and other objects of the invention are accomplished by providing a system and process for inputting client transaction data for a particular industry, organizing the client transactional data into a three dimensional array, generating statistical values for each entity-criterion over time by measuring (v), (n) and (d), weighting each statistical value and summing all values to form a single entity-criterion score, weighting each entity-criterion score and summing all entity-criterion scores to form a single entity score, comparing the entity score to a pre-determined threshold, 15 and displaying the entities so that those engaging in fraudulent behavior are easily identified.

20

25 The system and process also include providing a list of actions for the human analyst to take against those entities with scores indicating fraudulent behavior, and monitoring the behavior of those entities to see if they change thereby providing some evidence that the entities were initially engaging in fraudulent behavior, as well as 30 providing information to more precisely adjust the weighting and threshold mechanism used by the system.

With these and other objectives, advantages and features of the invention that may become apparent, the

nature of the invention may be more clearly understood by reference to the following detailed description of the invention, the appended claims, and to the several drawings attached herein.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a hardware block diagram of the client data input apparatus of the present invention;

10

Fig. 2 is a hardware block diagram of the pre-processor apparatus of the present invention;

Fig. 3 is a hardware block diagram of the monitoring and tracking apparatus of the present invention;

15

Fig. 4 is a block flow diagram showing modules and steps for generating the Value (v) statistic;

20

Fig. 4A is a block flow diagram showing modules and steps for performing the sums and squares calculation used in generating the Value (v) statistic;

Fig. 5 is a block flow diagram showing modules and steps for generating the Normalcy (n) statistic;

25

Fig. 6 is a block flow diagram showing modules and steps for generating the Change (c) statistic;

30

Fig. 7 is a block flow diagram showing modules and steps for a preferred embodiment of the present invention; and

Fig. 8 is a model showing the theoretical data structure used by the present invention.

DETAILED DESCRIPTION OF THE INVENTION

5 Referring now to the drawings wherein like elements refer to like reference numerals throughout, Figs. 1, 2 and 3 are hardware block diagrams for a fraud pre-processor detection system according to the present invention. An objective of the present invention is to provide a
10 mechanism for analyzing transaction data in filtering and targeting potentially fraudulent activities of individual entities. Entities may include health care providers, bank account holders, or other individuals to be analyzed for fraud risk "targeting". The term targeting refers to
15 isolating from a universe of entities a sub-set of those entities most likely to be engaging in fraudulent activities.

20 Specifically, Fig. 1 illustrates the client data input apparatus for inputting and organizing large amounts of client transactional data. Fig. 2 illustrates the pre-processor apparatus for evaluating an entity by generating statistical values for each of a number of pre-selected criteria, weighting and summing each
25 statistical value generated to form a single entity-criterion score, weighting and summing each entity-criterion score to form a single entity score, and comparing this entity score to a predetermined threshold. Fig. 3 illustrates the monitoring and tracking apparatus
30 for displaying a "risk list" of entities with indicators as to their level of risk with respect to engaging in fraudulent activities, generating a list of actions to take against those entities with the highest risk, monitoring

5 those entities against which actions have been taken in
 order to determine if behavior has been modified, and
 providing data for further refining weighting and threshold
 values to more accurately identify fraud in light of such
modified behavior.

10 A preferred embodiment of the data processed by
 this invention relates to the medical industry wherein
 transactional data (e.g., billing records or insurance
 claims) for a particular entity within a group (e.g., a
 cardio-thoracic surgeon within a group of surgeons or
 doctors, or a periodontist within a group of dentists, and
 so on) is analyzed to determine if patterns of fraudulent
 behavior can be detected. However, the processes and
 15 apparatus described herein can be applied to any other
 client transactional data for any number of industries.

20 It has been found in the preferred embodiment, that
 by examining client transactional data according to certain
 criteria, patterns can be detected. In the health care
 administration process, data collection occurs in a variety
 of ways: claims processing, subscriber enrollment,
 provider certification, reimbursement-accounting, and
 investigations. In some cases, the data is collected from
 25 external sources (e.g., provider certification). In other
 cases, the data is derived from internal processes (e.g.,
 reimbursement-accounting). In health care, most of the
 claims/subscriber/provider related data occurs as a normal
 course of doing business. Also included as "data" is that
 30 information developed as a result of internal or external
 investigation regarding a particular case.

Administrators of large plans may involve millions of subscribers and hundreds of thousands of providers. Consequently, there is a tremendous amount of data, and this aids the perpetrator of fraud or abuse since, with 5 some care, he or she can lie buried and virtually undetected. Because of the economics of the industry, most data processing systems are usually optimized for the rapid processing of claims, thereby limiting the kind of ad hoc analysis that is helpful in combating fraud and abuse.

10

As a result of the huge amount of client transactional data available for analysis, it is necessary to store in the criteria memory 16 (Fig. 1) a number of analysis criteria for use in the system as indicators of fraud. It has been found that for any domain environment subject to fraud there are many observable criteria that can be found in transaction data and history records. By understanding the operational characteristics of fraud in a particular domain, it is possible to select specific 15 criteria than can be used to detect the risk of fraud. 20 These criteria are determined manually through the user's knowledge of the fraud environment (e.g., how physicians commit fraud) and how the related behavior is represented in the transaction data and history. The objective of 25 selecting the criteria is to identify independent variables that can reflect behaviors that are focused on maximizing income and minimizing investment.

In general, these criteria can include two types: 30 (1) information which can be summed; or (2) information which can be counted. In the medical industry, examples of these criteria can include the number of claims filed by a provider (counted), the total amount billed (summed), the

number of critical versus non-critical patients (counted), the number of given patients seen by a doctor over the period of a day, week or year (counted), and so forth. These criteria can also include specific patterns 5 indicating fraud such as billing charges for intensive care ambulance services for a non-critical patient. These specific patterns, however, address only a small number of the possible ways fraud or abuse can be committed, and are therefore of limited use. Consequently, the more general 10 criteria lending themselves to broader analysis are preferred, such as those listed previously.

Typically, these criteria are selected according to their potential for detecting occurrences of fraud or 15 tracking the specific behavior of an entity under analysis, usually by the standards for a particular industry as established by past experience, or by an expert in intelligence gathering and analysis. In essence, the selection of these criteria acts as one step in the 20 filtering process whereby essential data from the large amounts of client transactional data is identified for use in the present invention. This criteria selection is analogous to a method used in military intelligence known 25 as "indications and warning" that monitored a variety of criteria for indications of increasing threat. When such a threat was detected, an alarm was triggered to warn the analyst. This same technique is used in the preferred embodiment to scan through the behavior history of all providers or patients and identify those having the highest 30 risk for fraud or abuse.

In the preferred embodiment, these analysis criteria are pre-programmed into the system. The analysis

5 criteria is used to identify the data elements contained in the data derived from the transaction process data structure received from the client data source 12. The selected data elements are included in a periodic extract process (e.g., monthly) which produces an extract data file to be processed by the pre-processor 110.

10 As shown in Fig. 1, once the criteria are client data converter 14 must then convert the client transaction data received from the client data source 12 for use in the pre-processor 110. Specifically, the client data converter 14 may be implemented as a hardware device or software program (such as a commercial database engine) 15 which accepts the extract data file and transforms it physically and organizationally in preparation for its use by the pre-processor 110.

20 The extract data file, typically an American Standard Computer Information Interface (ASCII) file or fixed record length file, is converted to a pre-processor input data file using the data structure required by the pre-processor 110. Some of the data elements received from the client data source 12 are directly used as criteria, 25 others must be converted before they are used. For instance, in the preferred embodiment, a criterion might be the number of claims submitted as a direct input data set. Another criterion might also use the ratio of claims per patient as an input set after dividing the number of claims submitted by the number of patients treated.

30 In addition to converting the data physically, the client data converter 14 also organizes the data in the

order expected by the pre-processor 110. Since peer-level comparison is one of the concepts used by the pre-processor 110, the data must be arranged into groups of like entities. In the preferred embodiment, this would mean 5 sorting the data into physician specialty groups, in the case of banking, this would involve sorting by like account/branch combinations. The objective is to put the data into groups of entities that can be expected to exhibit common behavior based upon their functional 10 characteristics (e.g., all cardiologists treat heart disease).

The data elements selected according to the analysis criteria can be arranged in any order except for 15 the master criteria. The master criteria is the criteria that will have the highest consistency for reflecting an entity's performance. For instance, good master criteria for health care would be either the number of claims submitted, or the number of patients treated -- since either is required for the physician to have any activity 20 during the month. The number of referrals, for instance, would be a poor master criteria since referrals are not required for the physician to operate.

25 Within each group, entity data is further organized according to time units. The pre-processor 110 can accept various time units (e.g., days, months, years, etc.) based upon the operational characteristics of the domain and the dynamics of the fraud threat. In the preferred embodiment, 30 it is usually adequate to analyze a one year historical sample of data composed of twelve monthly time units. In the case of bank fraud, since the dynamics are much

greater, the system would analyze ninety, one day time units.

The organization of the pre-processor input data file can be theoretically envisioned as a three-dimensional cube or array (as shown in Fig. 8,) where the x-axis contains the entities under analysis, the y-axis contains the criteria that are to be measured in determining the entity's performance, and the z-axis represents time divided into time units. Accordingly, each entity within a group can be evaluated according to each criterion over a specified amount of time. This creates the concept of an entity-criterion "bucket," wherein statistical scores or values can be generated for each entity according to each criterion.

Once the data are in a form capable of being stored in an electronic database 18, they can be connected to an appropriate controller 10 which is adapted to coordinate, control, monitor and execute commands necessary for operating the client data input apparatus 100, the pre-processor apparatus 110, as well as the monitoring and tracking apparatus 120, as shown in Figs. 1, 2 and 3, respectively. The functions performed by controller 10 are communicated to the hardware elements of Figs. 1, 2 and 3 via various control lines running from the controller 10 to its associated equipment of Figs. 1, 2 and 3. The controller 10 can comprise any appropriate microprocessor or minicomputer control device including, but not limited to, a PC-based controller or a dedicated control machine. In addition, the controller can comprise software implemented as a control program.

As shown in Fig. 2, once the pre-processor input data file is constructed and stored in the client database 18, it can be fed into the statistical analysis engine 50 of the pre-processor apparatus 110. The input file is fed 5 to the statistical analysis engine 50 as demonstrated by the following sequence: group 1, criterion 1, entity 1...n; criterion 2, entity 1...n; criterion n, entity 1...n; group 2...n, etc.. An analysis is performed on each entity-criterion combination by determining three 10 types of statistical measures: Value (v), Normalcy (n) and Change (d). These measures are a key component in the reliability and accuracy of identifying fraudulent behavior. They are designed to detect critical characteristics related to fraudulent behavior.

15 The first statistical measure is generated by the value processor 22. The value processor 22 generates a 20 Value (v) which determines "how important" the entity is within the group. This is a sort of "damage potential" measurement where the entity could be highly suspect in behavior but have an impact in dollars so small that he or she is not worth going after. The measure is simply the 25 average of the total (t) observations for an entity, divided by the average of the total observations for the criterion among all the entities in the group, times one hundred. That is, calculate the mean for the criterion 30 values for an entity by: (1) determining the number of active time units (Nt); (2) finding the sum of the values for all active time units (Sty); and (3) dividing Sty by Nt to find the mean of all time units for the entity (Mty).

When the Mty for each entity is calculated these values are summed for the group ($S(Mty)$) and then divided by the number of entities in the group (Ne) to find the mean criteria value for the group (Mc). Each entity's Mty is then divided by Mc , and the result is multiplied by 100 to find the (v) . The weight (e.g., multiplier) for the value measure for this criteria is looked up in the weight table 46 and then applied to (v) to find the weighted value score (sv). This (sv) is stored in the statistical value memory 28. (Figs. 4 and 4A give detailed block flow diagrams of the processes by which the weighted value (sv) score is generated.)

The second statistical measure is generated by the normalcy processor 24. Normalcy (n) is a measure of how much the entity conforms to the normal behavior of the group for this particular criterion. The measure is important because it is rarely possible that any particular entity can know how his or her behavior appears within the behavior of the total group.

The normalcy (n) measure is the statistical "standard score" for the entity. Calculating a standard score is relatively simple and well known in the art, and therefore any routine software implementation will suffice. Once the standard score for the entity is calculated, the weight for the normalcy measure for this criterion is looked up in the weight table 46 and then applied to (n) to find the weighted normalcy score (sn). This (sn) is stored in the statistical value memory 28. (Fig. 5 gives a detailed block flow diagram of the process by which the weighted normalcy (sn) score is generated.)

5 The third statistical measures is generated by the
change processor 26. Change (d) is measured by performing
a standard single linear regression on the observation (t)
values for the entity, estimating the value for the last t,
10 comparing the estimated to the actual value for the last t,
multiplying the difference by a constant and determining if
the actual value is greater or less than the estimate by
more than the allowed amount. As with the generation of
15 the Normalcy (n) measure, the calculation is relatively
simple and well known in the art, and therefore any routine
software implementation for a standard single linear
regression will suffice. Once the standard single linear
regression for the entity is calculated, the weight for the
change measure for this criteria is looked up in the weight
15 table 46 and then applied to (d) to find the weighted
change score (sd). This (sd) is stored in the statistical
value memory 28. (Fig. 6 gives a detailed block flow
diagram of the process by which the weighted change (sd)
score is generated.)

20

20 Returning to Fig. 2, the entity-criterion score
generator 30 generates a single score for each
entity-criterion. The entity-criterion score generator 30
retrieves the (sv), (sn) and (sd) scores for each
25 entity-criterion and sums them to form a single
entity-criterion score. The weight for each
entity-criterion score for each entity is looked up in the
weight table 46 and is then applied to the entity-criterion
score to find the weighted entity-criterion score which is
30 stored in the entity-criterion memory 32.

The entity score generator 34 takes each of the
weighted entity-criterion scores from the entity-criterion

memory 32 and sums all the weighted entity-criterion scores to form a single entity score. This score is then stored in the entity score database 36.

5 The risk analysis processor 38 takes the single entity score for all entities processed and compares them to a threshold value derived from the threshold table 48. The entities are then given a label corresponding to their amount of risk in relation to the threshold value.

10

As shown in Fig. 3, the entities are displayed by the risklist display 40, which uses text, tables, graphs, and graphical indicators to identify the levels of risk for each entity with respect to their engagement in fraudulent activities. In the preferred embodiment, entities are identified with a "red ball" to indicate high risk, a "yellow ball" to indicate medium risk, and a "green ball" to indicate low risk. This risklist of entities may be provided to an appropriate printer/plotter device 54 or display 52 to show the degree of risk for each entity.

15 One object of the present invention is to provide information to a human analyst in order to assist the analyst in making decisions and taking actions on appropriate risk entity targets. The risk action generator 25 42 offers a list of actions that can be taken against an entity. In the preferred embodiment, these actions can include a notice to the provider of performance aberration, or a request for an explanation of the behavior in a particular area. Actions are carefully defined to be 30 implementable in enterprise operations and traceable within the system. A generic list of actions would include: monitor, notify, warn, threaten, and cancel.

The analyst user of the fraud pre-processor detection system scans the results of the analysis from the risklist display 40 and selects entities that appear to exhibit the highest risk of fraud. The analyst decides 5 whether an action should be taken against an entity. The purpose of taking action is twofold: (1) to induce a change in the entity's behavior that is observable by the system and which will indicate his or her prior fraudulent actions (by highlighting those behavior changes in 10 subsequent analysis); and (2) to modify charges in such a way that savings produced by the fraud pre-processor detection system can be detected and commented upon. The system can also provide support to the conduct of investigations as a result of the observations and 15 conclusions that are developed.

If an action is selected by the analyst user, the actions taken against a particular entity are recorded and monitored by the risk behavior monitor 44. The risk 20 behavior monitor 44 stores the record of the actions, annotations that may be recorded by the analyst, the identity of the analyst, and other information important to maintaining knowledge of the actions related to each entity in the entity status database 47.

25 The summary and savings generator 45 provides automatic review of the results of the system on a monthly, or other selected period, basis. The summary and savings generator 45 summarizes the number of entities reviewed in 30 each category along with the total amount of dollars involved. Entities identified in various levels of risk are summarized, along with the number and type of actions taken. The system tracks actions and measures taken prior

to action having been taken, and subsequent performance for entities after actions have been taken. The summary and savings generator 45 provides a unique apparatus and process for determining the two most critical factors in 5 anti-fraud management: (1) specific aberrant behavior detection; and (2) cost reduction/savings.

The summary and savings generator 45 is connected to a group summary processor 49 and an individual summary 10 processor 51. The group summary processor 49 performs a monthly summary of the statistics relating to the group under review, such as the size of the group, the number of high risk entities, amount charged by the group, amount charged by the high risk entities in dollars and percent of 15 the total charges, total actions taken by type (e.g., level 1, 2, 3, etc.), the total amount claimed for savings, and so forth.

The individual summary processor 51 works only on 20 those entities that have been identified with an action flag in the entity status database 47 through the risk action monitor 44 in order to perform specific aberrant behavior detection. When the user analyst initiates one of 25 the actions, the level of the action, the date taken, the date to initiate subsequent review and the evaluation period to review, are all recorded in the entity status database 47.

The individual summary processor 51 operates on a 30 general assumption that most individuals who are knowingly doing something wrong, will, when attention is called to this fact by an external source, modify their behavior to reduce their risk of discovery and/or punitive action.

Consequently, the apparatus and process used in the invention focuses on behavior modification and the subsequent detection of the changes making the assumption that there is a high degree of correlation between the 5 observed changes and the areas of the entity's conduct that were either fraudulent or, at least, suspect.

Individual entities are screened and scored for their "risk" regarding fraud potentials during an initial 10 pass, or first review pass, through the system. The user analyst performs a review of the output scores in the measured criteria generated by the risklist display 40 and determines that some action generated by the risk action generator 42 should be taken against an entity. He uses 15 the risk action monitor 44 to record the action taken. The entity's record is attached with the appropriate action flag, along with the dates of action and the dates for subsequent review and added to the entity status database 47. The normal period of delay from action to subsequent 20 review is 90 days.

The actions generated by the risk action generator 42 are generally a range of options such as: (1) monitor 25 (i.e., place behavior under continuing review without notifying the individual); (2) notify (i.e., communicate with the individual that general behavior patterns have been detected that appear to be in question - but provide only limited details); (3) warn (i.e., communicate to the individual that specific patterns 30 of questionable behavior have been determined, indicating the nature of the behavior and suggesting self review); (4) threaten investigation (i.e., communicate that unacceptable behavior has been detected and that unless corrected

immediately detailed investigation will result); (5) initiate investigation (*i.e.*, communicate that unacceptable behavior of such dimension has been detected that an investigation is being opened and the individual can expect 5 additional action within a specified period of time); and (6) place on probation (*i.e.*, unless detected unacceptable behavior is changed within a specified time, the individual will be dropped as a provider of services/products).

10 While the exact wording or communications will be unique to each client application, this generic range of actions will be used and the risk action monitor 44 will record each action implemented in the entity status database 47. Since the actions are coded and graded for 15 increasing degrees of severity, the subsequent monitoring of individual performance will take into account the severity of the action when scoring response behavior. For example, a large response (*i.e.*, change in behavior) from a relatively low level action (e.g., #2 - Notify) would be 20 scored higher than a low response for the same action, or for a higher severity action.

25 The individual summary processor 51 determines which entities to review from the entity status database 47. It also determines from the date of the action and the delay period when a second pass review of the entities should be performed. Once an entity is selected for a second pass review, the individual summary processor 51 initiates the system to analyze client transaction data 30 from the client data source 12 from the date the action was taken to the date when the second pass review was ordered, for that particular entity. The individual summary processor 51 automatically retrieves the prior data on the

entity saved from the first pass review along with the results of the second pass review. The separate passes are extracted from the 36 entity score database as distinct performance periods on either side of the action-taken date. The data for the initial performance period and the subsequent performance period are compared for differences. Downward changes for charges are recorded as potential savings resulting from the action taken. The changes in specific criteria of measurement (e.g., in the preferred embodiment, the number of patients seen; the number of claims submitted; particular procedures rendered) are detected and used as an indicator of potential areas of fraud in the initial (i.e., first pass review) behavior period. Since the severity of the action taken will reflect on the degree of change to be expected, a multiplier from the 46 weight table is applied to amplify the differences observed according to the action taken. Also, since the prior performance of the entity with respect to his or her change trend will bear on the validity of the difference findings, a trend factor from the trend factor table 53 is applied. If the entity's prior performance is highly variable, the validity of the difference findings is reduced. Conversely, if the prior behavior trend is found to be very stable, significant differences will be amplified.

The results of the behavior change analysis performed by the individual summary processor 51 are presented to the analyst. Where the difference results indicate a high degree of correlation with the risk criteria identified in the first review pass, the analyst can assume a causal relationship with the action and that the areas that have changed are suggestive of aberrant

behavior prior to the action. The analyst can then use the behavior changes as a means of reviewing highly specific areas of the entity's prior performance, as well as modify the weight and threshold values found in the 46 weight table and 48 threshold table to better refine the detection process for subsequent reviews. The technique embodied in the invention provides several unique and previously unavailable capabilities to the process of fraud detection.

As shown in Figs. 4, 4A, 5 and 6, details of the block flow diagrams showing the steps for generating the statistical measures for Value (v), Normalcy (n) and Change (d) are shown. As the means for generating these statistical measures are well-known in the art, they will not be described in detail herein. It should be noted, however, that although the means for generating these statistical measures are well-known, the application of these statistical measures to transactional data in order to analyze such data for purposes of fraud detection is considered a novel feature of the invention.

As shown in Fig. 7, which details a block flow diagram showing the steps for performing the pre-processor method, the initial stages of the process serve to filter and highlight suspect individuals, as described previously. Once suspect individuals are detected, prioritized and highlighted by the system an analyst will review the results and actions generated by the risklist display 40 and risk action generator 42, respectively, to determine what action to take. This automated filtering and prioritization process is critical because reviewing all of the behavior of all entities in sufficient detail to precisely identify fraud is impractical. The amount of

data normally collected in the transaction-billing process is far to great to permit manual, or human review. And the potential variations for fraud in such industries as health care are far too many, and much too flexible, to build a 5 completely automated pattern detection system that would be reliable for any protracted period. Consequently, the present invention reduces the focus of the human analyst to the most likely payoff.

10 The combined statistical approach of the pre-processor apparatus 110 highlights the most critical areas of behavior in performing the filtering, that is: value, normalcy, and change. These measures, when applied to carefully selected behavior criteria (i.e., selected for 15 their potential to exhibit profit-taking) produce a high degree of validity in showing the analyst where to look. The output of the first pass review is normally sufficient for the analyst to decide to take action.

20 This element of the present invention is one of the unique components of the process - the concept of introducing process control theory to the detection of fraud in large transaction data histories. There has not, heretofore, been available a methodology that was 25 implementable in an automated process that could record behavior changes in specific entities and determine the savings resulting from the counter-fraud program. As a result of the first pass review, organizations currently suffering from high losses due to fraud have a means of not 30 only identifying candidate targets, but they can modify the behavior of those targets in order to obtain savings and further focus on specific fraudulent behavior. The results of the process can offer the analyst with the option of

declaring savings or using the highlighted areas of change as a means of initiating deeper investigation into very specific areas of the entity's performance.

5 While the initial use of the process is focused on fraud in the health care industry, it is apparent that the technique described here can also be used to detect fraud and identify resulting savings in a variety of environments (e.g., banking, mortgage, worker's compensation, etc.).

10

15

20

25

30

I CLAIM:

1. A fraud detection system for targeting a potentially fraudulent entity within a particular industry,
5 comprising:

statistical analysis engine for generating statistical values for said entity according to at least one of a plurality of analysis criteria;

10 entity-criterion score generator for weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

15 entity score generator for weighting and summing each said entity-criterion scores to form a first entity score;

20 computer memory for storing said statistical values, said entity-criterion scores, and said first entity score; and

25 risk analysis processor for comparing said first entity score to a predetermined threshold, whereby the result of said comparison indicates whether said entity is engaging in fraudulent activity.

30 2. The system according to claim 1, further comprising controller means for coordinating, controlling, monitoring and executing commands necessary for operating said fraud detection system.

3. The system according to claim 1, further comprising a database for storing said first entity score.

4. The system according to claim 1, wherein said 5 statistical values include a value statistic (v), an normalcy statistic (n), and a change statistic (d) for each of said criteria.

5. The system according to claim 1, further 10 comprising client data converter for converting client records containing data elements useful in measuring the performance of said entity to a format compatible with said fraud detection system.

15 6. The system according to claim 1, further comprising a three dimensional array for organizing and storing said entity and said criteria according to predefined time units, said array comprising an x, y and z axis, said y-axis containing said entity, said x-axis 20 containing said criteria, and said z-axis containing said time units.

7. The system according to claim 6, wherein said 25 array creates entity-criterion buckets wherein said statistical values for each criterion can be generated for said entity within said time period.

8. The system according to claim 1, further 30 comprising a risklist display for displaying in graphical form whether said entity is engaging in fraudulent behavior.

9. The system according to claim 1, further comprising:

5 risk action generator for displaying possible actions to initiate against said entity, and selecting one of said actions to modify said entity's behavior;

10 risk behavior monitor for storing a record of said selected action taken against said entity; and

15 summary and savings generator for identifying the amount of money gained or lost by said action.

10. The system according to claim 9, wherein said summary and savings means include:

15 a second entity score presenting said entity's behavior after said action has occurred;

20 individual summary processor for comparing said first entity score with said second entity score, measuring the degree of change therefrom, weighting the degree of change using a severity factor and a trend factor, and displaying a result indicating whether said action decreased the fraudulent activity of said entity;

25 group summary processor for generating group statistics for a plurality of entities.

30 11. The system according to claim 10, wherein said individual summary processor includes adjuster for adjusting said weighting values for said criteria scoring means and said entity scoring means, and said threshold

values for said risk analysis means, in accordance with said group statistics from said summary and savings means.

12. A fraud detection system for targeting
5 potentially fraudulent entities within a particular
industry, comprising:

10 client data source for inputting client records
containing data elements useful in measuring the
performance of an entity by evaluating a plurality of
criteria over a set of time period;

15 there dimensional array for organizing and storing
said entity, said criteria, and said time period,
comprising an x, y and z axis, said y-axis containing said
entity, said x-axis containing said criteria, and said
z-axis containing at least one of said time units;

20 statistical analysis engine for generating a value
statistic (v), a normalcy statistic (n), and a change
statistic (d) for each of said criteria;

25 entity-criterion score generator for weighting and
summing each of said statistical values thereby forming an
entity-criterion score for each of said criteria;

entity score generator for weighting and summing
each of said entity-criterion scores to form an entity
score for said entity;

30

risk analysis processor for comparing said entity
score with pre-defined threshold values; and

risklist display for displaying in a graphical manner whether said entity is engaging in fraudulent behavior.

5 13. A computer program for targeting potentially fraudulent entities within a particular industry, said computer program comprising the steps of:

10 inputting client records containing data elements useful in measuring the performance of an entity by evaluating a plurality of criteria over a set time period;

15 organizing and storing said entity, said criteria, and said time period, into a three dimensional array comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing at least one of said time units;

20 generating a value statistic (v), a normalcy statistic (n), and a change statistic (d) for each of said criteria;

25 weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said criteria;

 weighting and summing each of said entity-criterion scores to form an entity score for said entity;

30 comparing said entity score with pre-defined threshold values; and

displaying graphically whether said entity is engaging in fraudulent behavior.

14. The computer program according to claim 13, 5 said computer program further comprising the step of encoding said computer program on a computer-readable medium.

15. A method of targeting potentially fraudulent 10 entities within a particular industry, comprising the steps of:

generating statistical values for entity according to a plurality of analysis criteria;

15 weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

20 weighing and summing each of said entity-criterion scores to form an entity score; and

comparing said entity score to a predetermined threshold.

25 16. The computer program according to claim 15, said computer program further comprising the step of encoding said computer program on a computer-readable medium.

30 17. A fraud pre-processor system for filtering and highlighting those individuals most likely to be engaging in fraudulent behavior, comprising:

a controller;

a computer program executed by said controller comprising:

5

client data converter for converting client data to format compatible with said fraud pre-processor system;

10

three dimensional array for storing a plurality of entities, analysis criteria, and time units such that for each of said entities statistical values for each of said criteria within each of said time units can be generated, weighted and summed to form an entity-criterion score;

15

entity-criterion score generator for generating said entity-criterion score;

20

entity score generator for weighting and summing each entity-criterion score for each entity to form an entity score;

25

risk analysis processor for comparing said entity score for each of said entities with a threshold value;

30

risklist display for displaying the results from said comparison in a manner identifying those entities most likely engaging in fraudulent behavior; and

35

a database for permanently storing, and a computer memory for temporarily storing, said analysis criteria, said statistical values, said entity-criterion score, said entity-score, said threshold values, and said results.

18. A fraud detection system, comprising statistical analysis engine for generating statistics for a plurality of entities according to a plurality of analysis criteria and time periods; and

5

pre-processor for evaluating said statistics to determine if said entities are engaging in fraudulent behavior.

10

19. The system according to claim 1, wherein said statistical analysis engine further comprises:

value processor for generating said value statistic (v) which determines the damage potential for said entity;

15

normalcy processor for generating said normalcy statistic (n) for measuring normal behavior for said entities; and change processor for generating said change statistic (d).

20

20. A fraud detection system, comprising:

risk analysis processor for comparing an entity score for an entity to a predetermined threshold;

25

risk action generator for displaying and selecting action to be taken said entity;

30

risk action method for storing a record of said actions; summary and savings generator for identifying the amount of money gained or lost by said action; and

adjustor for adjusting weighting values used to generate said entity score, and said threshold values for said risk analysis means, in accordance with said statistics from said summary and savings generator, in order to better estimate if said entity is engaging in fraudulent behavior.

10

15

20

25

30

1/9

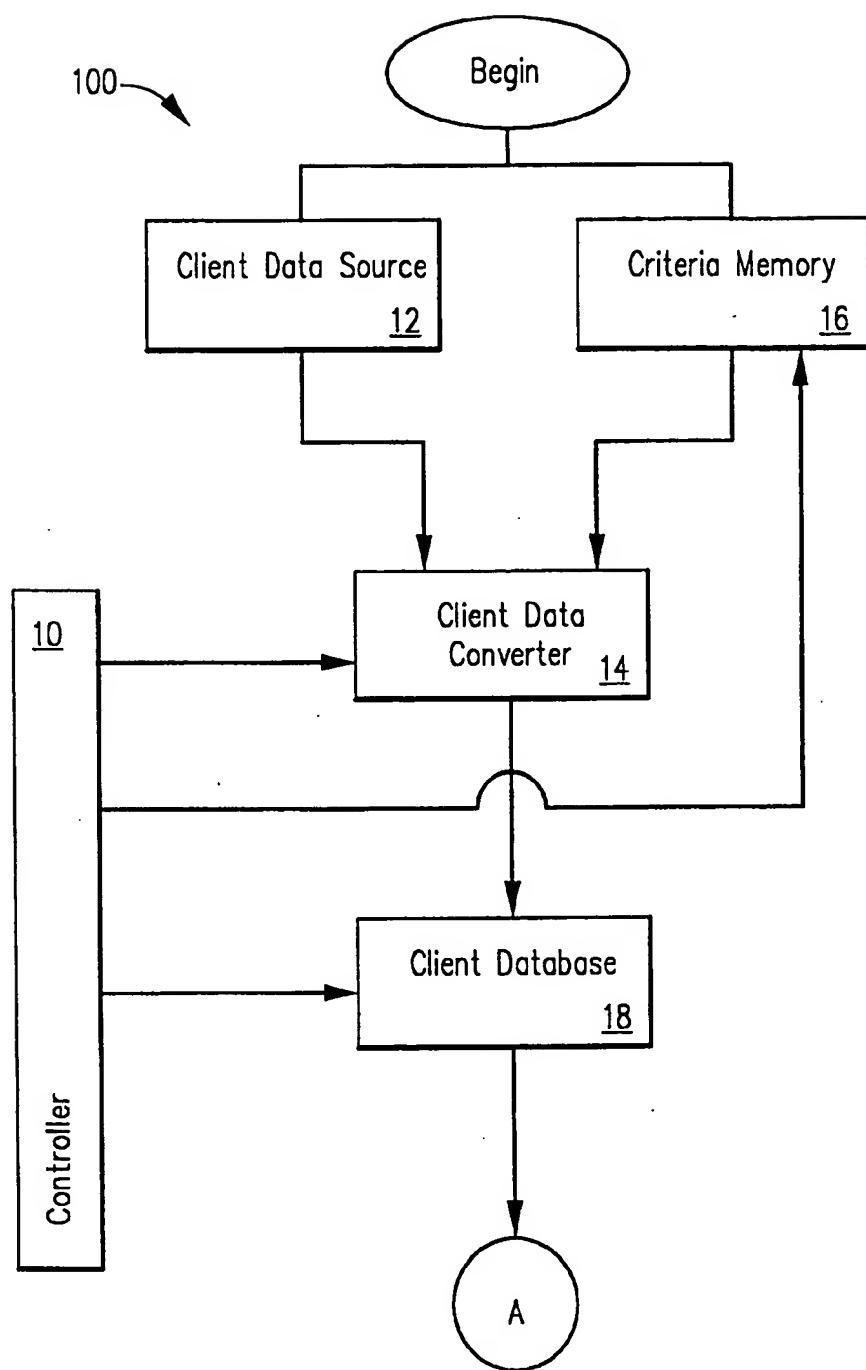
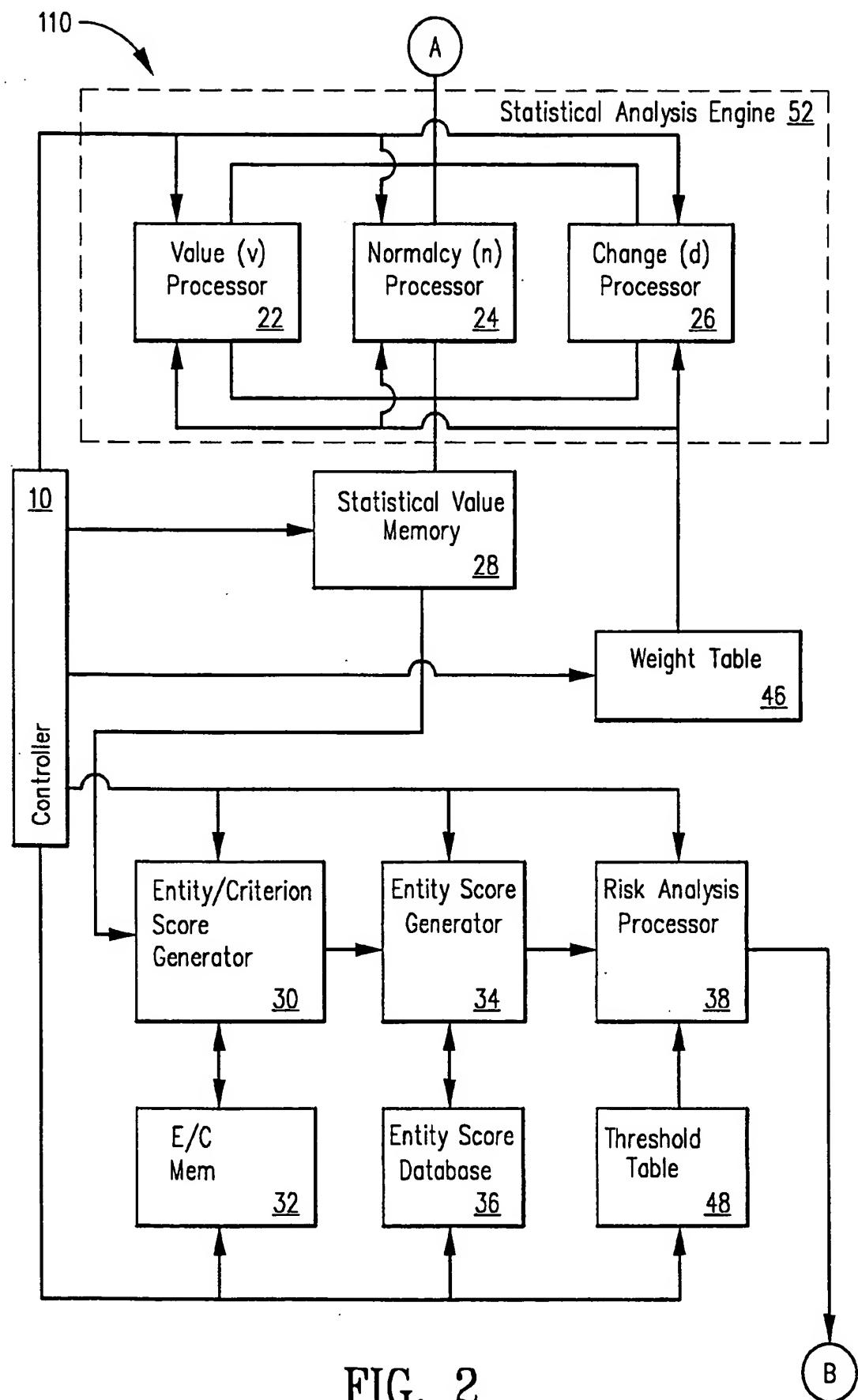


FIG. 1

2/9



3/9

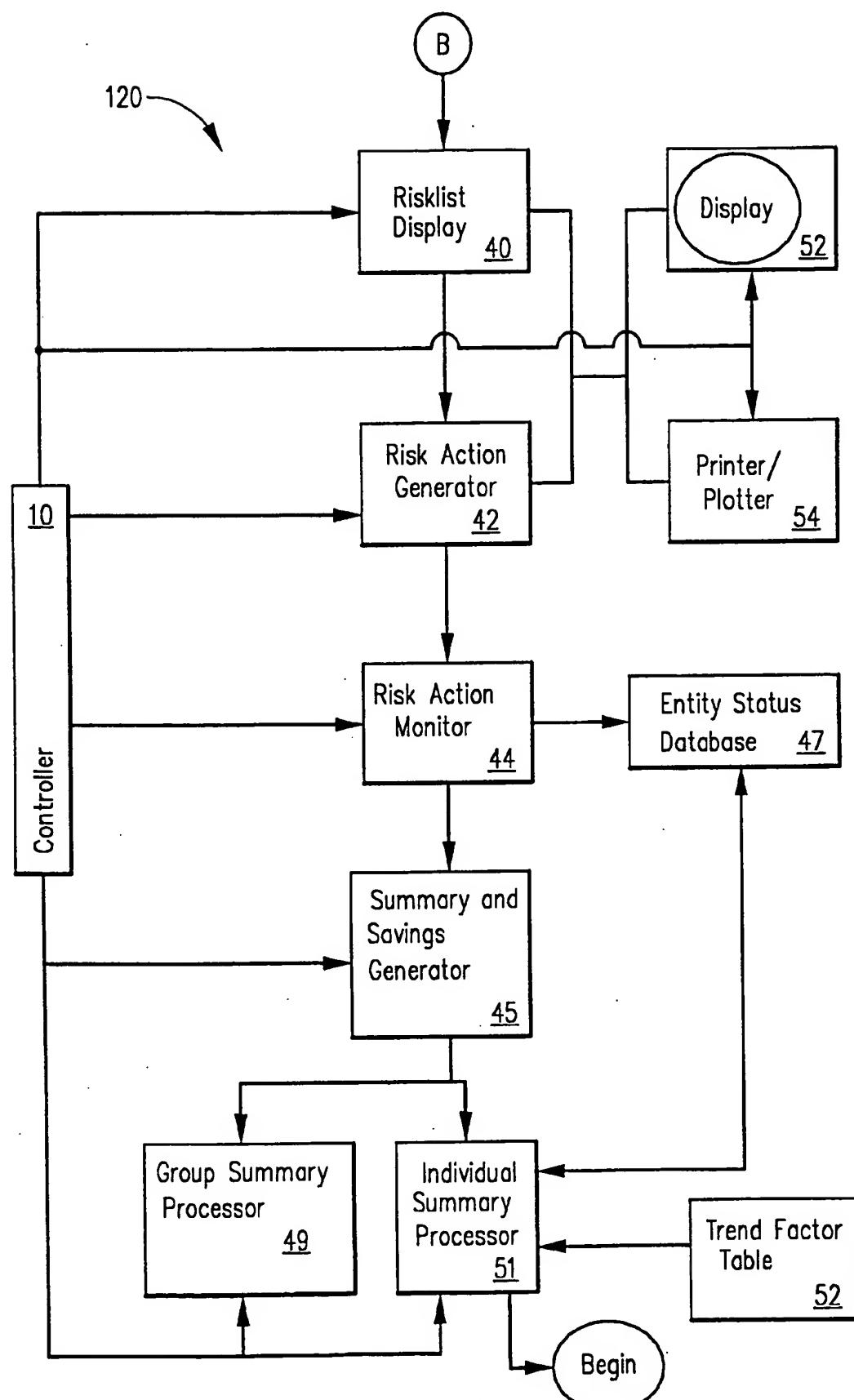
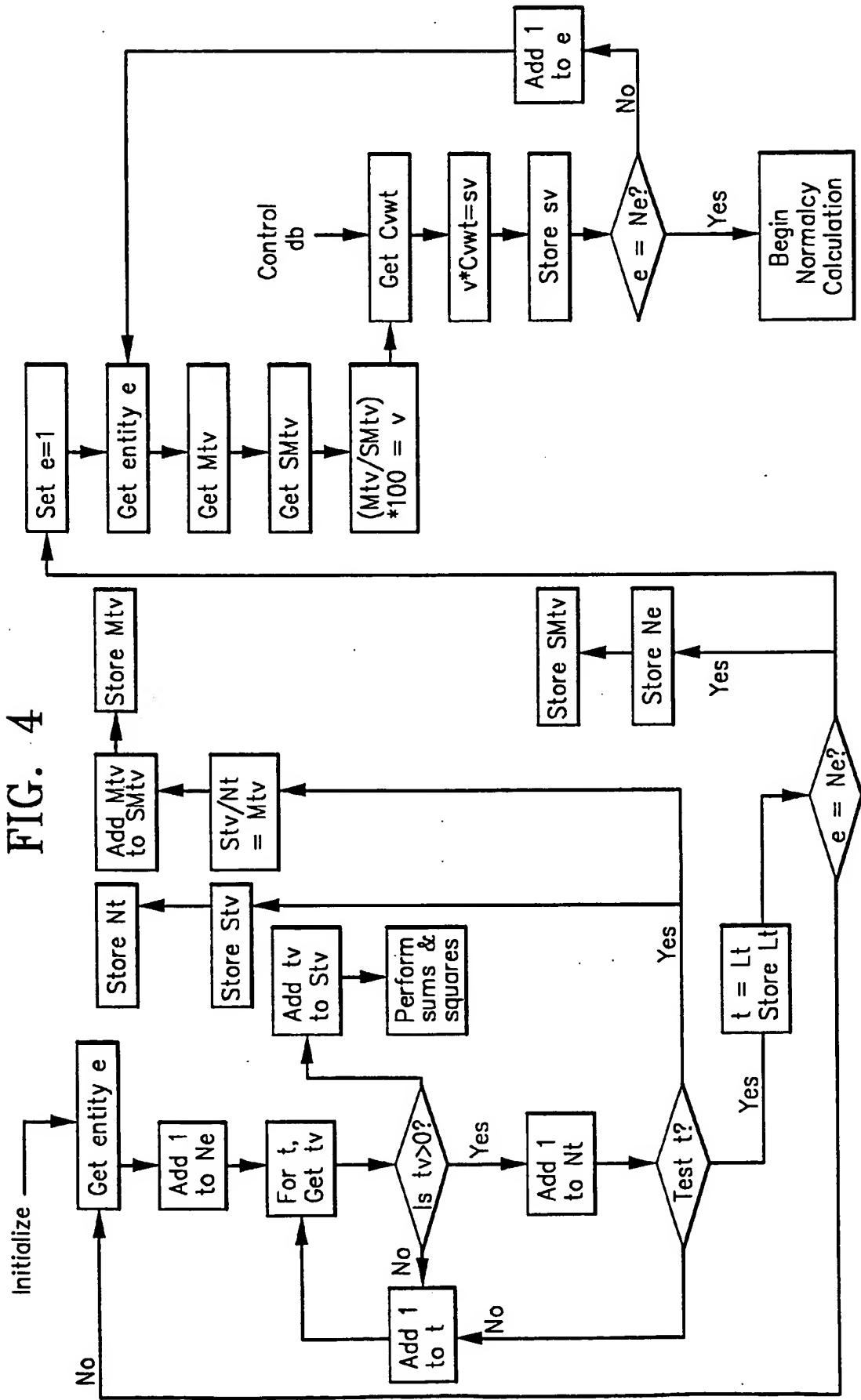


FIG. 3

4/9



5/9

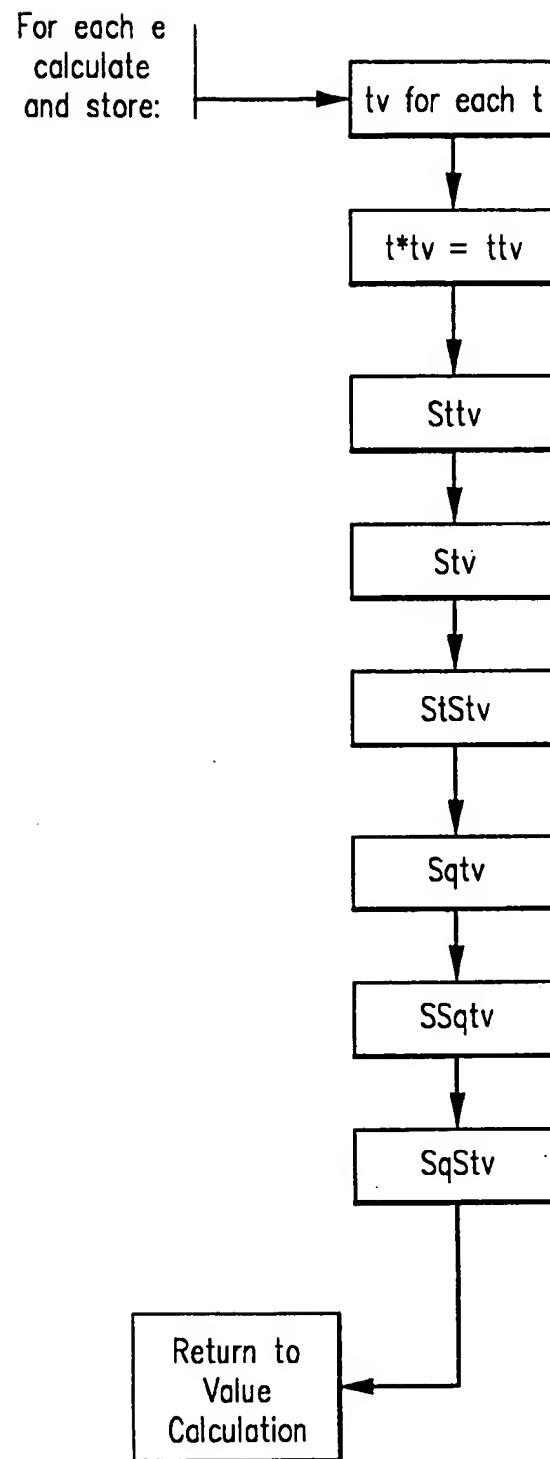
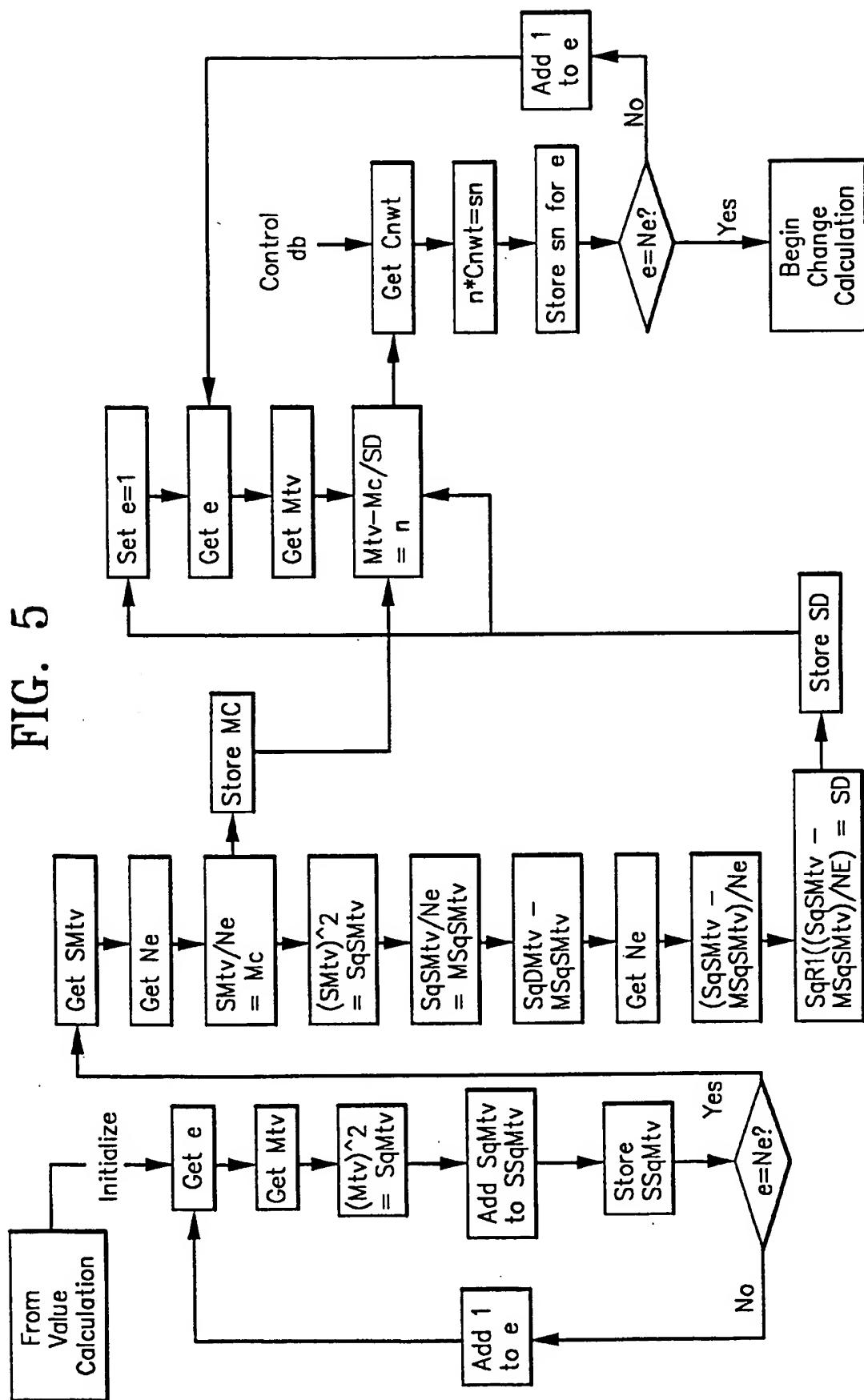


FIG. 4A

6 / 9

FIG. 5



7/9

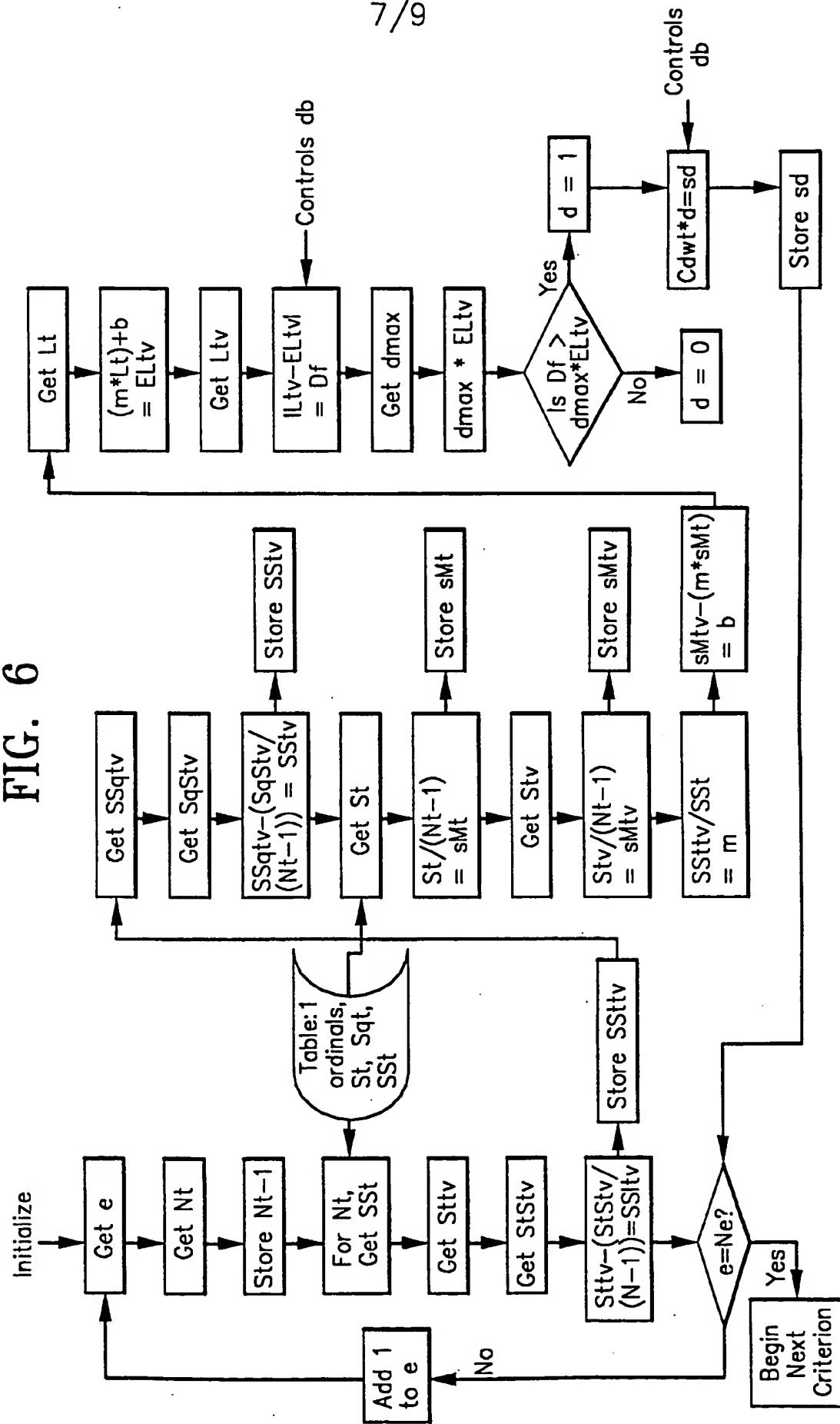


FIG. 6

80/9

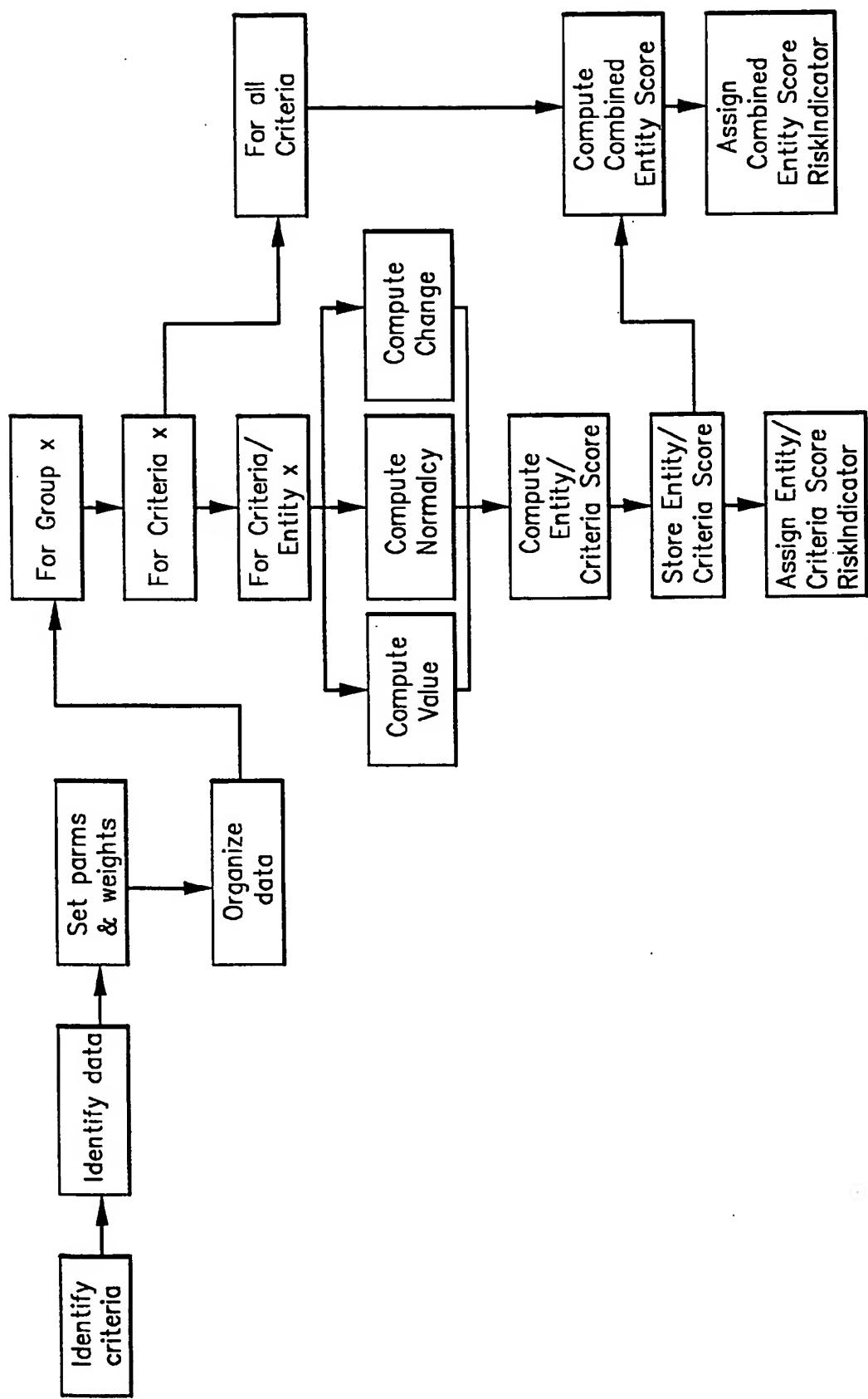


FIG. 7

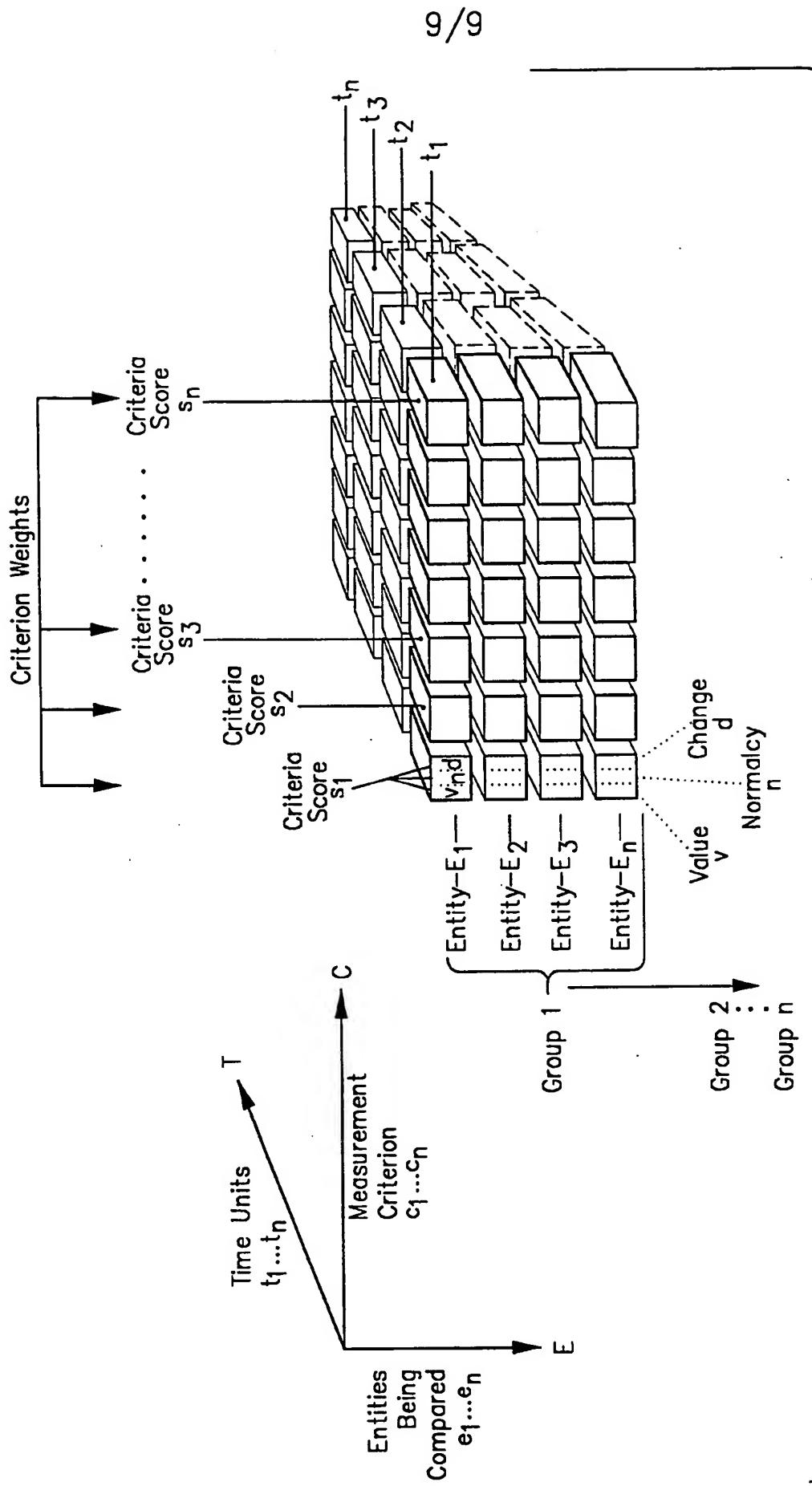


FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/10352

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G06F 15/00, 17/60; G06G 7/52
US CL :364/401M, 401R, 406, 408

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 364/401M, 401R, 406, 408

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y, E	US, A, 5,557,514 (SEARE ET AL) 17 September 1996, cols. 4, 19, 21	1-20
Y	US, A, 5,253,164 (HOLLOWAY ET AL) 12 October 1993, col. 3	1-20
Y	US, A, 5,359,509 (LITTLE ET AL) 25 October 1994, cols. 4,5,8,9,13,14	1-20

Further documents are listed in the continuation of Box C.

See patent family annex.

• Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
• "A" document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
• "E" earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
• "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
• "O" document referring to an oral disclosure, use, exhibition or other means		
• "P" document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
22 OCTOBER 1996

Date of mailing of the international search report

27 NOV 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer

GAIL O. HAYES

Telephone No. (703) 305-9711

Joni Bell

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US96/10352

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, DIALOG, STN

Search Terms: track? or monitor? or detect?; fraud? or aberration? or deviat?; historic? or past or previous; analys? or credit? or check?



AE

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : G06F 15/00, 17/60, G06G 7/52		A1	(11) International Publication Number: WO 97/00483 (43) International Publication Date: 3 January 1997 (03.01.97)
<p>(21) International Application Number: PCT/US96/10352</p> <p>(22) International Filing Date: 14 June 1996 (14.06.96)</p> <p>(30) Priority Data: 08/490,984 15 June 1995 (15.06.95) US</p> <p>(71) Applicant: FRAUDetect, L.L.C. [US/US]; Suite 312, 10400 Eaton Place, Fairfax, VA 22030 (US).</p> <p>(72) Inventor: COFOD, Robert, K.; 282 Prince's Lane, Harwood, MD 20776 (US).</p> <p>(74) Agent: GROSSMAN, Jon, D.; 2101 L Street, N.W., Washington, DC 20037 (US).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p>Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>	
<p>(54) Title: PROCESS AND APPARATUS FOR DETECTING FRAUD</p> <pre> graph TD A[Identify criteria] --> B[Identify data] B --> C[Set norms & weights] C --> D[Organize data] D --> E[For Group x] E --> F[For Criteria x] F --> G[For Criteria/Entity x] G --> H[Compute Value] G --> I[Compute Normalcy] G --> J[Compute Change] H --> K[Compute Entity/Criteria Score] I --> K J --> K K --> L[Store Entity/Criteria Score] L --> M[Compute Combined Entity Score] M --> N[Assign Combined Entity Score - RiskIndicator] </pre>			
<p>(57) Abstract</p> <p>A system and method for processing transaction data in filtering and targeting potentially fraudulent entities within a particular industry relies upon client data input apparatus for inputting client/provider/entity data records containing data elements for measuring an entity's performance according to pre-selected criteria, where entity data, criteria data and time data are organized and stored in three-dimensional array. A pre-processor (110) generates a score for each entity-criterion (1..n) by generating a value statistic (v), a normalcy statistic (n), and a change statistic (d), weighting each statistical value, and summing the values. An entity scoring generator (30) generates a single entity score for each entity by weighting each entity-criterion score and summing the entity-criterion scores. The entity scores are compared to pre-defined threshold values for detecting those entities most likely to be engaging in fraudulent activity and a graphical display entities according to level of risk.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

PROCESS AND APPARATUS FOR DETECTING FRAUD

5

FIELD OF THE INVENTION

10

The invention relates to a process and apparatus for analyzing client data against pre-selected criteria over a pre-determined period of time in order to identify those individuals or entities engaging in activity indicative of potential fraudulent behavior within a particular industry.

BACKGROUND OF THE ART

15

20

25

Many industries, such as medical and banking, suffer enormous losses due to the theft of money or services through fraudulent means, such as submitting insurance claims for unseen patients or forging checks. In the medical industry, for example, it is estimated that approximately \$100 billion of U.S. annual health care expenditures is stolen through fraud or abuse. Even more alarming is that some experts believe this figure only represents the portion of which we are aware. The actual figure may be significantly higher, perhaps as much as twice the amount.

30

The problems surrounding fraud detection are many. Since most companies maintain relatively few investigators, limited resources are available to perform the analysis and decision-making necessary to accurately attack the problem and ensure measurable results. In addition, the possibility of wrongly accusing a client of fraudulent behavior adds an extra burden of caution and proof. Confidentiality, government involvement, market

competition, and judicial processes add other limits to the options for attacking the problem.

Other problems in fraud detection include highly varying types of targets, possible scenarios, and innovative fraud techniques requiring heavy reliance on skilled human analysts and decision makers to detect, recognize, define and initiate actions. In addition, each industry, such as the aforementioned medical and banking industries, generates a massive and dynamic quantity of information from which relevant "target" information must be extracted. Furthermore, there is a time sensitivity factor in processing and acting upon newly detected targets or instances of fraud.

15

Due to the above-cited difficulties in detecting those persons engaging in fraudulent activities, billions of dollars worth of stolen money or services have largely been ignored or passed on as a cost of operation.

20

Most cases of fraud are discovered as a result of tip-offs or indications from industry program integrity reviews. These discoveries have often been cited as being only the tip of the iceberg, yet most investigative staffs are already inundated with the cases on hand.

25

The problem is not unlike another well-known problem in military intelligence and targeting circles. It is known as a "target rich environment," where the number of possible targets far exceeds the resources available for attack. The military problem, like the one faced in combating health care and banking fraud and abuse, is fundamentally one of resource management. The problem

requires a method or system by which an industry may increase ensurance that the resources available to it are being spent on targets with the highest potential payback. Consequently, there is a need for processing, ordering and analyzing the volumes of transactional data, such as 5 billing information and insurance claims, associated with a particular entity, such as a physician working within the health care industry, and extrapolating from such information whether the client is engaging in fraudulent 10 behavior. This need for specifically targeting perpetrators of fraud or abuse becomes more readily apparent in light of the fact that there may be tens or hundreds of thousands of clients per industry. Accordingly, the present invention answers that specific 15 need.

SUMMARY OF THE INVENTION

It is therefore apparent from the above that there exists a significant need for a system and process which 20 effectively uses transactional data in a manner that identifies potentially fraudulent behavior and areas of greatest need for investigation. It is therefore one of the objectives of this invention to provide a fraud pre-processor detection system. Fraud can be detected for 25 many industries dealing with a large number of clients, such as the aforementioned medical and banking industries.

Another object of this invention is to provide a process for analyzing large volumes of transaction data for 30 the purpose of detecting indications of fraud.

Another object of this invention is to provide a system and process which automatically prioritizes those

clients or entities which are most likely to be engaging in fraudulent behavior.

5 Another object of this invention is to provide a system and process for establishing thresholds indicating fraud.

10 An additional object of this invention is to provide a system and process for filtering, processing and displaying data in a manner where it may be used as an 10 analysis tool to support the human decision making process.

15 Another object of this invention is to provide a system and process for organizing data in a three dimensional array consisting of pre-selected criteria along the x-axis, groups of entities along the y-axis, and time units along the z-axis, thereby allowing entity-criterion analysis; that is analyzing each entity within a group (a single chiropractor out of a group of chiropractors) 20 according to each criterion (e.g., the number of patients seen by the chiropractor, or the amount of claims submitted to the insurance carrier), over a pre-selected period of time units (e.g., days, months or years).

25 Yet a further object of this invention is to provide a system and process for generating statistical values for each entity-criterion to measure Value (v), or the relative importance of the entity with respect to the entire group, Normalcy (n), or how the entity conforms to 30 the normal behavior of the group for a particular criterion, and Change (d), or the amount of change in activity by the entity over a period of time.

5 It is also an objective of this invention to weight the statistical values for Value (v), Normalcy (n) and Change (d) according to the relative importance of each value within a criterion with respect to targeting entities engaging in fraudulent activity.

10 Yet a further object of this invention is to add all the statistical values for each entity-criterion together to create a single entity-criterion score, weight the score according to its fraud targeting potential, sum all the individual entity-criterion scores to form a final entity score, and compare that score to a threshold value to determine the risk of fraudulent behavior.

15 A further object of this invention is to display all the entities under analysis in a manner that allows easy identification of those entities engaging in behavior most likely indicative of fraud.

20 Another object of the invention is to provide a list of actions to take against those entities suspected of fraudulent behavior so that a human analyst may select the appropriate action.

25 It is a further object of this invention to provide a system and process for monitoring those entities suspected of engaging in fraud once an action has been taken against the entity, and determining whether the behavior of the entity changes, and measuring such change, 30 thereby providing some evidence that fraudulent behavior was accurately identified and corrected while also providing a mechanism by which the weighting of the relative importance of various criteria, entity-criterion

scores and entity scores can be modified to more accurately detect the presence of fraudulent behavior.

5 It is yet a further object of this invention to provide a system and process for generating a change meter which can be translated into dollar amounts to calculate savings to a particular industry.

10 Briefly described, these and other objects of the invention are accomplished by providing a system and process for inputting client transaction data for a particular industry, organizing the client transactional data into a three dimensional array, generating statistical values for each entity-criterion over time by measuring 15 (v), (n) and (d), weighting each statistical value and summing all values to form a single entity-criterion score, weighting each entity-criterion score and summing all entity-criterion scores to form a single entity score, comparing the entity score to a pre-determined threshold, 20 and displaying the entities so that those engaging in fraudulent behavior are easily identified.

25 The system and process also include providing a list of actions for the human analyst to take against those entities with scores indicating fraudulent behavior, and monitoring the behavior of those entities to see if they change thereby providing some evidence that the entities were initially engaging in fraudulent behavior, as well as 30 providing information to more precisely adjust the weighting and threshold mechanism used by the system.

With these and other objectives, advantages and features of the invention that may become apparent, the

nature of the invention may be more clearly understood by reference to the following detailed description of the invention, the appended claims, and to the several drawings attached herein.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a hardware block diagram of the client data input apparatus of the present invention;

10

Fig. 2 is a hardware block diagram of the pre-processor apparatus of the present invention;

Fig. 3 is a hardware block diagram of the monitoring and tracking apparatus of the present invention;

15

Fig. 4 is a block flow diagram showing modules and steps for generating the Value (v) statistic;

20

Fig. 4A is a block flow diagram showing modules and steps for performing the sums and squares calculation used in generating the Value (v) statistic;

Fig. 5 is a block flow diagram showing modules and steps for generating the Normalcy (n) statistic;

25

Fig. 6 is a block flow diagram showing modules and steps for generating the Change (c) statistic;

30

Fig. 7 is a block flow diagram showing modules and steps for a preferred embodiment of the present invention; and

Fig. 8 is a model showing the theoretical data structure used by the present invention.

DETAILED DESCRIPTION OF THE INVENTION

5 Referring now to the drawings wherein like elements refer to like reference numerals throughout, Figs. 1, 2 and 3 are hardware block diagrams for a fraud pre-processor detection system according to the present invention. An objective of the present invention is to provide a
10 mechanism for analyzing transaction data in filtering and targeting potentially fraudulent activities of individual entities. Entities may include health care providers, bank account holders, or other individuals to be analyzed for fraud risk "targeting". The term targeting refers to
15 isolating from a universe of entities a sub-set of those entities most likely to be engaging in fraudulent activities.

20 Specifically, Fig. 1 illustrates the client data input apparatus for inputting and organizing large amounts of client transactional data. Fig. 2 illustrates the pre-processor apparatus for evaluating an entity by generating statistical values for each of a number of pre-selected criteria, weighting and summing each
25 statistical value generated to form a single entity-criterion score, weighting and summing each entity-criterion score to form a single entity score, and comparing this entity score to a predetermined threshold. Fig. 3 illustrates the monitoring and tracking apparatus for displaying a "risk list" of entities with indicators as
30 to their level of risk with respect to engaging in fraudulent activities, generating a list of actions to take against those entities with the highest risk, monitoring

5 those entities against which actions have been taken in
 order to determine if behavior has been modified, and
 providing data for further refining weighting and threshold
 values to more accurately identify fraud in light of such
modified behavior.

10 A preferred embodiment of the data processed by
 this invention relates to the medical industry wherein
 transactional data (e.g., billing records or insurance
claims) for a particular entity within a group (e.g., a
15 cardio-thoracic surgeon within a group of surgeons or
doctors, or a periodontist within a group of dentists, and
so on) is analyzed to determine if patterns of fraudulent
behavior can be detected. However, the processes and
15 apparatus described herein can be applied to any other
client transactional data for any number of industries.

20 It has been found in the preferred embodiment, that
 by examining client transactional data according to certain
criteria, patterns can be detected. In the health care
administration process, data collection occurs in a variety
of ways: claims processing, subscriber enrollment,
provider certification, reimbursement-accounting, and
investigations. In some cases, the data is collected from
25 external sources (e.g., provider certification). In other
cases, the data is derived from internal processes (e.g.,
reimbursement-accounting). In health care, most of the
claims/subscriber/provider related data occurs as a normal
course of doing business. Also included as "data" is that
30 information developed as a result of internal or external
investigation regarding a particular case.

Administrators of large plans may involve millions of subscribers and hundreds of thousands of providers. Consequently, there is a tremendous amount of data, and this aids the perpetrator of fraud or abuse since, with 5 some care, he or she can lie buried and virtually undetected. Because of the economics of the industry, most data processing systems are usually optimized for the rapid processing of claims, thereby limiting the kind of ad hoc analysis that is helpful in combating fraud and abuse.

10

As a result of the huge amount of client transactional data available for analysis, it is necessary to store in the criteria memory 16 (Fig. 1) a number of analysis criteria for use in the system as indicators of 15 fraud. It has been found that for any domain environment subject to fraud there are many observable criteria that can be found in transaction data and history records. By understanding the operational characteristics of fraud in a particular domain, it is possible to select specific 20 criteria than can be used to detect the risk of fraud. These criteria are determined manually through the user's knowledge of the fraud environment (e.g., how physicians commit fraud) and how the related behavior is represented in the transaction data and history. The objective of 25 selecting the criteria is to identify independent variables that can reflect behaviors that are focused on maximizing income and minimizing investment.

In general, these criteria can include two types: 30 (1) information which can be summed; or (2) information which can be counted. In the medical industry, examples of these criteria can include the number of claims filed by a provider (counted), the total amount billed (summed), the

number of critical versus non-critical patients (counted), the number of given patients seen by a doctor over the period of a day, week or year (counted), and so forth. These criteria can also include specific patterns 5 indicating fraud such as billing charges for intensive care ambulance services for a non-critical patient. These specific patterns, however, address only a small number of the possible ways fraud or abuse can be committed, and are therefore of limited use. Consequently, the more general 10 criteria lending themselves to broader analysis are preferred, such as those listed previously.

Typically, these criteria are selected according to their potential for detecting occurrences of fraud or 15 tracking the specific behavior of an entity under analysis, usually by the standards for a particular industry as established by past experience, or by an expert in intelligence gathering and analysis. In essence, the selection of these criteria acts as one step in the 20 filtering process whereby essential data from the large amounts of client transactional data is identified for use in the present invention. This criteria selection is analogous to a method used in military intelligence known 25 as "indications and warning" that monitored a variety of criteria for indications of increasing threat. When such a threat was detected, an alarm was triggered to warn the analyst. This same technique is used in the preferred embodiment to scan through the behavior history of all providers or patients and identify those having the highest 30 risk for fraud or abuse.

In the preferred embodiment, these analysis criteria are pre-programmed into the system. The analysis

5 criteria is used to identify the data elements contained in the data derived from the transaction process data structure received from the client data source 12. The selected data elements are included in a periodic extract process (e.g., monthly) which produces an extract data file to be processed by the pre-processor 110.

10 As shown in Fig. 1, once the criteria are established and programmed into the criteria memory 16, the client data converter 14 must then convert the client transaction data received from the client data source 12 for use in the pre-processor 110. Specifically, the client data converter 14 may be implemented as a hardware device or software program (such as a commercial database engine) 15 which accepts the extract data file and transforms it physically and organizationally in preparation for its use by the pre-processor 110.

20 The extract data file, typically an American Standard Computer Information Interface (ASCII) file or fixed record length file, is converted to a pre-processor input data file using the data structure required by the pre-processor 110. Some of the data elements received from the client data source 12 are directly used as criteria, 25 others must be converted before they are used. For instance, in the preferred embodiment, a criterion might be the number of claims submitted as a direct input data set. Another criterion might also use the ratio of claims per patient as an input set after dividing the number of claims submitted by the number of patients treated.

30 In addition to converting the data physically, the client data converter 14 also organizes the data in the

order expected by the pre-processor 110. Since peer-level comparison is one of the concepts used by the pre-processor 110, the data must be arranged into groups of like entities. In the preferred embodiment, this would mean 5 sorting the data into physician specialty groups, in the case of banking, this would involve sorting by like account/branch combinations. The objective is to put the data into groups of entities that can be expected to exhibit common behavior based upon their functional 10 characteristics (e.g., all cardiologists treat heart disease).

The data elements selected according to the analysis criteria can be arranged in any order except for 15 the master criteria. The master criteria is the criteria that will have the highest consistency for reflecting an entity's performance. For instance, good master criteria for health care would be either the number of claims submitted, or the number of patients treated -- since either is required for the physician to have any activity 20 during the month. The number of referrals, for instance, would be a poor master criteria since referrals are not required for the physician to operate.

25 Within each group, entity data is further organized according to time units. The pre-processor 110 can accept various time units (e.g., days, months, years, etc.) based upon the operational characteristics of the domain and the dynamics of the fraud threat. In the preferred embodiment, 30 it is usually adequate to analyze a one year historical sample of data composed of twelve monthly time units. In the case of bank fraud, since the dynamics are much

greater, the system would analyze ninety, one day time units.

5 The organization of the pre-processor input data file can be theoretically envisioned as a three-dimensional cube or array (as shown in Fig. 8,) where the x-axis contains the entities under analysis, the y-axis contains the criteria that are to be measured in determining the entity's performance, and the z-axis represents time
10 divided into time units. Accordingly, each entity within a group can be evaluated according to each criterion over a specified amount of time. This creates the concept of an entity-criterion "bucket," wherein statistical scores or values can be generated for each entity according to each
15 criterion.

20 Once the data are in a form capable of being stored in an electronic database 18, they can be connected to an appropriate controller 10 which is adapted to coordinate, control, monitor and execute commands necessary for operating the client data input apparatus 100, the pre-processor apparatus 110, as well as the monitoring and tracking apparatus 120, as shown in Figs. 1, 2 and 3, respectively. The functions performed by controller 10 are
25 communicated to the hardware elements of Figs. 1, 2 and 3 via various control lines running from the controller 10 to its associated equipment of Figs. 1, 2 and 3. The controller 10 can comprise any appropriate microprocessor or minicomputer control device including, but not limited
30 to, a PC-based controller or a dedicated control machine. In addition, the controller can comprise software implemented as a control program.

As shown in Fig. 2, once the pre-processor input data file is constructed and stored in the client database 18, it can be fed into the statistical analysis engine 50 of the pre-processor apparatus 110. The input file is fed 5 to the statistical analysis engine 50 as demonstrated by the following sequence: group 1, criterion 1, entity 1...n; criterion 2, entity 1...n; criterion n, entity 1...n; group 2...n, etc.. An analysis is performed on each entity-criterion combination by determining three 10 types of statistical measures: Value (v), Normalcy (n) and Change (d). These measures are a key component in the reliability and accuracy of identifying fraudulent behavior. They are designed to detect critical characteristics related to fraudulent behavior.

15

The first statistical measure is generated by the value processor 22. The value processor 22 generates a Value (v) which determines "how important" the entity is within the group. This is a sort of "damage potential" 20 measurement where the entity could be highly suspect in behavior but have an impact in dollars so small that he or she is not worth going after. The measure is simply the average of the total (t) observations for an entity, divided by the average of the total observations for the 25 criterion among all the entities in the group, times one hundred. That is, calculate the mean for the criterion values for an entity by: (1) determining the number of active time units (Nt); (2) finding the sum of the values for all active time units (Sty); and (3) dividing Sty by Nt to find the mean of all time units for the entity (Mty). 30

When the Mty for each entity is calculated these values are summed for the group ($S(Mty)$) and then divided by the number of entities in the group (N_e) to find the mean criteria value for the group (M_c). Each entity's Mty is then divided by M_c , and the result is multiplied by 100 to find the (v). The weight (e.g., multiplier) for the value measure for this criteria is looked up in the weight table 46 and then applied to (v) to find the weighted value score (sv). This (sv) is stored in the statistical value memory 28. (Figs. 4 and 4A give detailed block flow diagrams of the processes by which the weighted value (sv) score is generated.)

The second statistical measure is generated by the normalcy processor 24. Normalcy (n) is a measure of how much the entity conforms to the normal behavior of the group for this particular criterion. The measure is important because it is rarely possible that any particular entity can know how his or her behavior appears within the behavior of the total group.

The normalcy (n) measure is the statistical "standard score" for the entity. Calculating a standard score is relatively simple and well known in the art, and therefore any routine software implementation will suffice. Once the standard score for the entity is calculated, the weight for the normalcy measure for this criterion is looked up in the weight table 46 and then applied to (n) to find the weighted normalcy score (sn). This (sn) is stored in the statistical value memory 28. (Fig. 5 gives a detailed block flow diagram of the process by which the weighted normalcy (sn) score is generated.)

The third statistical measures is generated by the change processor 26. Change (d) is measured by performing a standard single linear regression on the observation (t) values for the entity, estimating the value for the last t, 5 comparing the estimated to the actual value for the last t, multiplying the difference by a constant and determining if the actual value is greater or less than the estimate by more than the allowed amount. As with the generation of the Normalcy (n) measure, the calculation is relatively 10 simple and well known in the art, and therefore any routine software implementation for a standard single linear regression will suffice. Once the standard single linear regression for the entity is calculated, the weight for the change measure for this criteria is looked up in the weight 15 table 46 and then applied to (d) to find the weighted change score (sd). This (sd) is stored in the statistical value memory 28. (Fig. 6 gives a detailed block flow diagram of the process by which the weighted change (sd) score is generated.)

20
25
30

Returning to Fig. 2, the entity-criterion score generator 30 generates a single score for each entity-criterion. The entity-criterion score generator 30 retrieves the (sv), (sn) and (sd) scores for each entity-criterion and sums them to form a single entity-criterion score. The weight for each entity-criterion score for each entity is looked up in the weight table 46 and is then applied to the entity-criterion score to find the weighted entity-criterion score which is stored in the entity-criterion memory 32.

35
40
45
50
55
60
65
70
75
80
85
90
95
100
105
110
115
120
125
130
135
140
145
150
155
160
165
170
175
180
185
190
195
200
205
210
215
220
225
230
235
240
245
250
255
260
265
270
275
280
285
290
295
300
305
310
315
320
325
330
335
340
345
350
355
360
365
370
375
380
385
390
395
400
405
410
415
420
425
430
435
440
445
450
455
460
465
470
475
480
485
490
495
500
505
510
515
520
525
530
535
540
545
550
555
560
565
570
575
580
585
590
595
600
605
610
615
620
625
630
635
640
645
650
655
660
665
670
675
680
685
690
695
700
705
710
715
720
725
730
735
740
745
750
755
760
765
770
775
780
785
790
795
800
805
810
815
820
825
830
835
840
845
850
855
860
865
870
875
880
885
890
895
900
905
910
915
920
925
930
935
940
945
950
955
960
965
970
975
980
985
990
995
1000
1005
1010
1015
1020
1025
1030
1035
1040
1045
1050
1055
1060
1065
1070
1075
1080
1085
1090
1095
1100
1105
1110
1115
1120
1125
1130
1135
1140
1145
1150
1155
1160
1165
1170
1175
1180
1185
1190
1195
1200
1205
1210
1215
1220
1225
1230
1235
1240
1245
1250
1255
1260
1265
1270
1275
1280
1285
1290
1295
1300
1305
1310
1315
1320
1325
1330
1335
1340
1345
1350
1355
1360
1365
1370
1375
1380
1385
1390
1395
1400
1405
1410
1415
1420
1425
1430
1435
1440
1445
1450
1455
1460
1465
1470
1475
1480
1485
1490
1495
1500
1505
1510
1515
1520
1525
1530
1535
1540
1545
1550
1555
1560
1565
1570
1575
1580
1585
1590
1595
1600
1605
1610
1615
1620
1625
1630
1635
1640
1645
1650
1655
1660
1665
1670
1675
1680
1685
1690
1695
1700
1705
1710
1715
1720
1725
1730
1735
1740
1745
1750
1755
1760
1765
1770
1775
1780
1785
1790
1795
1800
1805
1810
1815
1820
1825
1830
1835
1840
1845
1850
1855
1860
1865
1870
1875
1880
1885
1890
1895
1900
1905
1910
1915
1920
1925
1930
1935
1940
1945
1950
1955
1960
1965
1970
1975
1980
1985
1990
1995
2000
2005
2010
2015
2020
2025
2030
2035
2040
2045
2050
2055
2060
2065
2070
2075
2080
2085
2090
2095
2100
2105
2110
2115
2120
2125
2130
2135
2140
2145
2150
2155
2160
2165
2170
2175
2180
2185
2190
2195
2200
2205
2210
2215
2220
2225
2230
2235
2240
2245
2250
2255
2260
2265
2270
2275
2280
2285
2290
2295
2300
2305
2310
2315
2320
2325
2330
2335
2340
2345
2350
2355
2360
2365
2370
2375
2380
2385
2390
2395
2400
2405
2410
2415
2420
2425
2430
2435
2440
2445
2450
2455
2460
2465
2470
2475
2480
2485
2490
2495
2500
2505
2510
2515
2520
2525
2530
2535
2540
2545
2550
2555
2560
2565
2570
2575
2580
2585
2590
2595
2600
2605
2610
2615
2620
2625
2630
2635
2640
2645
2650
2655
2660
2665
2670
2675
2680
2685
2690
2695
2700
2705
2710
2715
2720
2725
2730
2735
2740
2745
2750
2755
2760
2765
2770
2775
2780
2785
2790
2795
2800
2805
2810
2815
2820
2825
2830
2835
2840
2845
2850
2855
2860
2865
2870
2875
2880
2885
2890
2895
2900
2905
2910
2915
2920
2925
2930
2935
2940
2945
2950
2955
2960
2965
2970
2975
2980
2985
2990
2995
3000
3005
3010
3015
3020
3025
3030
3035
3040
3045
3050
3055
3060
3065
3070
3075
3080
3085
3090
3095
3100
3105
3110
3115
3120
3125
3130
3135
3140
3145
3150
3155
3160
3165
3170
3175
3180
3185
3190
3195
3200
3205
3210
3215
3220
3225
3230
3235
3240
3245
3250
3255
3260
3265
3270
3275
3280
3285
3290
3295
3300
3305
3310
3315
3320
3325
3330
3335
3340
3345
3350
3355
3360
3365
3370
3375
3380
3385
3390
3395
3400
3405
3410
3415
3420
3425
3430
3435
3440
3445
3450
3455
3460
3465
3470
3475
3480
3485
3490
3495
3500
3505
3510
3515
3520
3525
3530
3535
3540
3545
3550
3555
3560
3565
3570
3575
3580
3585
3590
3595
3600
3605
3610
3615
3620
3625
3630
3635
3640
3645
3650
3655
3660
3665
3670
3675
3680
3685
3690
3695
3700
3705
3710
3715
3720
3725
3730
3735
3740
3745
3750
3755
3760
3765
3770
3775
3780
3785
3790
3795
3800
3805
3810
3815
3820
3825
3830
3835
3840
3845
3850
3855
3860
3865
3870
3875
3880
3885
3890
3895
3900
3905
3910
3915
3920
3925
3930
3935
3940
3945
3950
3955
3960
3965
3970
3975
3980
3985
3990
3995
4000
4005
4010
4015
4020
4025
4030
4035
4040
4045
4050
4055
4060
4065
4070
4075
4080
4085
4090
4095
4100
4105
4110
4115
4120
4125
4130
4135
4140
4145
4150
4155
4160
4165
4170
4175
4180
4185
4190
4195
4200
4205
4210
4215
4220
4225
4230
4235
4240
4245
4250
4255
4260
4265
4270
4275
4280
4285
4290
4295
4300
4305
4310
4315
4320
4325
4330
4335
4340
4345
4350
4355
4360
4365
4370
4375
4380
4385
4390
4395
4400
4405
4410
4415
4420
4425
4430
4435
4440
4445
4450
4455
4460
4465
4470
4475
4480
4485
4490
4495
4500
4505
4510
4515
4520
4525
4530
4535
4540
4545
4550
4555
4560
4565
4570
4575
4580
4585
4590
4595
4600
4605
4610
4615
4620
4625
4630
4635
4640
4645
4650
4655
4660
4665
4670
4675
4680
4685
4690
4695
4700
4705
4710
4715
4720
4725
4730
4735
4740
4745
4750
4755
4760
4765
4770
4775
4780
4785
4790
4795
4800
4805
4810
4815
4820
4825
4830
4835
4840
4845
4850
4855
4860
4865
4870
4875
4880
4885
4890
4895
4900
4905
4910
4915
4920
4925
4930
4935
4940
4945
4950
4955
4960
4965
4970
4975
4980
4985
4990
4995
5000
5005
5010
5015
5020
5025
5030
5035
5040
5045
5050
5055
5060
5065
5070
5075
5080
5085
5090
5095
5100
5105
5110
5115
5120
5125
5130
5135
5140
5145
5150
5155
5160
5165
5170
5175
5180
5185
5190
5195
5200
5205
5210
5215
5220
5225
5230
5235
5240
5245
5250
5255
5260
5265
5270
5275
5280
5285
5290
5295
5300
5305
5310
5315
5320
5325
5330
5335
5340
5345
5350
5355
5360
5365
5370
5375
5380
5385
5390
5395
5400
5405
5410
5415
5420
5425
5430
5435
5440
5445
5450
5455
5460
5465
5470
5475
5480
5485
5490
5495
5500
5505
5510
5515
5520
5525
5530
5535
5540
5545
5550
5555
5560
5565
5570
5575
5580
5585
5590
5595
5600
5605
5610
5615
5620
5625
5630
5635
5640
5645
5650
5655
5660
5665
5670
5675
5680
5685
5690
5695
5700
5705
5710
5715
5720
5725
5730
5735
5740
5745
5750
5755
5760
5765
5770
5775
5780
5785
5790
5795
5800
5805
5810
5815
5820
5825
5830
5835
5840
5845
5850
5855
5860
5865
5870
5875
5880
5885
5890
5895
5900
5905
5910
5915
5920
5925
5930
5935
5940
5945
5950
5955
5960
5965
5970
5975
5980
5985
5990
5995
6000
6005
6010
6015
6020
6025
6030
6035
6040
6045
6050
6055
6060
6065
6070
6075
6080
6085
6090
6095
6100
6105
6110
6115
6120
6125
6130
6135
6140
6145
6150
6155
6160
6165
6170
6175
6180
6185
6190
6195
6200
6205
6210
6215
6220
6225
6230
6235
6240
6245
6250
6255
6260
6265
6270
6275
6280
6285
6290
6295
6300
6305
6310
6315
6320
6325
6330
6335
6340
6345
6350
6355
6360
6365
6370
6375
6380
6385
6390
6395
6400
6405
6410
6415
6420
6425
6430
6435
6440
6445
6450
6455
6460
6465
6470
6475
6480
6485
6490
6495
6500
6505
6510
6515
6520
6525
6530
6535
6540
6545
6550
6555
6560
6565
6570
6575
6580
6585
6590
6595
6600
6605
6610
6615
6620
6625
6630
6635
6640
6645
6650
6655
6660
6665
6670
6675
6680
6685
6690
6695
6700
6705
6710
6715
6720
6725
6730
6735
6740
6745
6750
6755
6760
6765
6770
6775
6780
6785
6790
6795
6800
6805
6810
6815
6820
6825
6830
6835
6840
6845
6850
6855
6860
6865
6870
6875
6880
6885
6890
6895
6900
6905
6910
6915
6920
6925
6930
6935
6940
6945
6950
6955
6960
6965
6970
6975
6980
6985
6990
6995
7000
7005
7010
7015
7020
7025
7030
7035
7040
7045
7050
7055
7060
7065
7070
7075
7080
7085
7090
7095
7100
7105
7110
7115
7120
7125
7130
7135
7140
7145
7150
7155
7160
7165
7170
7175
7180
7185
7190
7195
7200
7205
7210
7215
7220
7225
7230
7235
7240
7245
7250
7255
7260
7265
7270
7275
7280
7285
7290
7295
7300
7305
7310
7315
7320
7325
7330
7335
7340
7345
7350
7355
7360
7365
7370
7375
7380
7385
7390
7395
7400
7405
7410
7415
7420
7425
7430
7435
7440
7445
7450
7455
7460
7465
7470
7475
7480
7485
7490
7495
7500
7505
7510
7515
7520
7525
7530
7535
7540
7545
7550
7555
7560
7565
7570
7575
7580
7585
7590
7595
7600
7605
7610
7615
7620
7625
7630
7635
7640
7645
7650
7655
7660
7665
7670
7675
7680
7685
7690
7695
7700
7705
7710
7715
7720
7725
7730
7735
7740
7745
7750
7755
7760
7765
7770
7775
7780
7785
7790
7795
7800
7805
7810
7815
7820
7825
7830
7835
7840
7845
7850
7855
7860
7865
7870
7875
7880
7885
7890
7895
7900
7905
7910
7915
7920
7925
7930
7935
7940
7945
7950
7955
7960
7965
7970
7975
7980
7985
7990
7995
8000
8005
8010
8015
8020
8025
8030
8035
8040
8045
8050
8055
8060
8065
8070
8075
8080
8085
8090
8095
8100
8105
8110
8115
8120
8125
8130
8135
8140
8145
8150
8155
8160
8165
8170
8175
8180
8185
8190
8195
8200
8205
8210
8215
8220
8225
8230
8235
8240
8245
8250
8255
8260
8265
8270
8275
8280
8285
8290
8295
8300
8305
8310
8315
8320
8325
8330
8335
8340
8345
8350
8355
8360
8365
8370
8375
8380
8385
8390
8395
8400
8405
8410
8415
8420
8425
8430
8435
8440
8445
8450
8455
8460
8465
8470
8475
8480
8485
8490
8495
8500
8505
8510
8515
8520
8525
8530
8535
8540
8545
8550
8555
8560
8565
8570
8575
8580
8585
8590
8595
8600
8605
8610
8615
8620
8625
8630
8635
8640
8645
8650
8655
8660
8665
8670
8675
8680
8685
8690
8695
8700
8705
8710
8715
8720
8725
8730
8735
8740
8745
8750
8755
8760
8765
8770
8775
8780
8785
8790
8795
8800
8805
8810
8815
8820
8825
8830
8835
8840
8845
8850
8855
8860
8865
8870
8875
8880
8885
8890
8895
8900
8905
8910
8915
8920
8925
8930
8935
8940
8945
8950
8955
8960
8965
8970
8975
8980
8985
8990
8995
9000
9005
9010
9015
9020
9025
9030
9035
9040
9045
9050
9055
9060
9065
9070
9075
9080
9085
9090
9095
9100
9105
9110
9115
9120
9125
9130
9135
9140
9145
9150
9155
9160
9165
9170
9175
9180
9185
9190
9195
9200
9205
9210
9215
9220
9225
9230
923

memory 32 and sums all the weighted entity-criterion scores to form a single entity score. This score is then stored in the entity score database 36.

5 The risk analysis processor 38 takes the single entity score for all entities processed and compares them to a threshold value derived from the threshold table 48. The entities are then given a label corresponding to their amount of risk in relation to the threshold value.

10 As shown in Fig. 3, the entities are displayed by the risklist display 40, which uses text, tables, graphs, and graphical indicators to identify the levels of risk for each entity with respect to their engagement in fraudulent 15 activities. In the preferred embodiment, entities are identified with a "red ball" to indicate high risk, a "yellow ball" to indicate medium risk, and a "green ball" to indicate low risk. This risklist of entities may be provided to an appropriate printer/plotter device 54 or 20 display 52 to show the degree of risk for each entity.

One object of the present invention is to provide 25 information to a human analyst in order to assist the analyst in making decisions and taking actions on appropriate risk entity targets. The risk action generator 42 offers a list of actions that can be taken against an entity. In the preferred embodiment, these actions can include a notice to the provider of performance aberration, or a request for an explanation of the behavior in a 30 particular area. Actions are carefully defined to be implementable in enterprise operations and traceable within the system. A generic list of actions would include: monitor, notify, warn, threaten, and cancel.

The analyst user of the fraud pre-processor detection system scans the results of the analysis from the risklist display 40 and selects entities that appear to exhibit the highest risk of fraud. The analyst decides 5 whether an action should be taken against an entity. The purpose of taking action is twofold: (1) to induce a change in the entity's behavior that is observable by the system and which will indicate his or her prior fraudulent actions (by highlighting those behavior changes in 10 subsequent analysis); and (2) to modify charges in such a way that savings produced by the fraud pre-processor detection system can be detected and commented upon. The system can also provide support to the conduct of 15 investigations as a result of the observations and conclusions that are developed.

If an action is selected by the analyst user, the actions taken against a particular entity are recorded and monitored by the risk behavior monitor 44. The risk 20 behavior monitor 44 stores the record of the actions, annotations that may be recorded by the analyst, the identity of the analyst, and other information important to maintaining knowledge of the actions related to each entity in the entity status database 47.

25 The summary and savings generator 45 provides automatic review of the results of the system on a monthly, or other selected period, basis. The summary and savings generator 45 summarizes the number of entities reviewed in 30 each category along with the total amount of dollars involved. Entities identified in various levels of risk are summarized, along with the number and type of actions taken. The system tracks actions and measures taken prior

to action having been taken, and subsequent performance for entities after actions have been taken. The summary and savings generator 45 provides a unique apparatus and process for determining the two most critical factors in 5 anti-fraud management: (1) specific aberrant behavior detection; and (2) cost reduction/savings.

10 The summary and savings generator 45 is connected to a group summary processor 49 and an individual summary processor 51. The group summary processor 49 performs a monthly summary of the statistics relating to the group under review, such as the size of the group, the number of high risk entities, amount charged by the group, amount charged by the high risk entities in dollars and percent of 15 the total charges, total actions taken by type (e.g., level 1, 2, 3, etc.), the total amount claimed for savings, and so forth.

20 The individual summary processor 51 works only on those entities that have been identified with an action flag in the entity status database 47 through the risk action monitor 44 in order to perform specific aberrant behavior detection. When the user analyst initiates one of 25 the actions, the level of the action, the date taken, the date to initiate subsequent review and the evaluation period to review, are all recorded in the entity status database 47.

30 The individual summary processor 51 operates on a general assumption that most individuals who are knowingly doing something wrong, will, when attention is called to this fact by an external source, modify their behavior to reduce their risk of discovery and/or punitive action.

Consequently, the apparatus and process used in the invention focuses on behavior modification and the subsequent detection of the changes making the assumption that there is a high degree of correlation between the 5 observed changes and the areas of the entity's conduct that were either fraudulent or, at least, suspect.

Individual entities are screened and scored for their "risk" regarding fraud potentials during an initial 10 pass, or first review pass, through the system. The user analyst performs a review of the output scores in the measured criteria generated by the risklist display 40 and determines that some action generated by the risk action generator 42 should be taken against an entity. He uses 15 the risk action monitor 44 to record the action taken. The entity's record is attached with the appropriate action flag, along with the dates of action and the dates for subsequent review and added to the entity status database 47. The normal period of delay from action to subsequent 20 review is 90 days.

The actions generated by the risk action generator 42 are generally a range of options such as: (1) monitor 25 (i.e., place behavior under continuing review without notifying the individual); (2) notify (i.e., communicate with the individual that general behavior patterns have been detected that appear to be in question - but provide only limited details); (3) warn (i.e., communicate to the individual that specific patterns 30 of questionable behavior have been determined, indicating the nature of the behavior and suggesting self review); (4) threaten investigation (i.e., communicate that unacceptable behavior has been detected and that unless corrected

immediately detailed investigation will result); (5) initiate investigation (*i.e.*, communicate that unacceptable behavior of such dimension has been detected that an investigation is being opened and the individual can expect 5 additional action within a specified period of time); and (6) place on probation (*i.e.*, unless detected unacceptable behavior is changed within a specified time, the individual will be dropped as a provider of services/products).

10 While the exact wording or communications will be unique to each client application, this generic range of actions will be used and the risk action monitor 44 will record each action implemented in the entity status database 47. Since the actions are coded and graded for 15 increasing degrees of severity, the subsequent monitoring of individual performance will take into account the severity of the action when scoring response behavior. For example, a large response (*i.e.*, change in behavior) from a relatively low level action (e.g., #2 - Notify) would be 20 scored higher than a low response for the same action, or for a higher severity action.

25 The individual summary processor 51 determines which entities to review from the entity status database 47. It also determines from the date of the action and the delay period when a second pass review of the entities should be performed. Once an entity is selected for a second pass review, the individual summary processor 51 initiates the system to analyze client transaction data 30 from the client data source 12 from the date the action was taken to the date when the second pass review was ordered, for that particular entity. The individual summary processor 51 automatically retrieves the prior data on the

entity saved from the first pass review along with the results of the second pass review. The separate passes are extracted from the 36 entity score database as distinct performance periods on either side of the action-taken date. The data for the initial performance period and the subsequent performance period are compared for differences. Downward changes for charges are recorded as potential savings resulting from the action taken. The changes in specific criteria of measurement (e.g., in the preferred embodiment, the number of patients seen; the number of claims submitted; particular procedures rendered) are detected and used as an indicator of potential areas of fraud in the initial (i.e., first pass review) behavior period. Since the severity of the action taken will reflect on the degree of change to be expected, a multiplier from the 46 weight table is applied to amplify the differences observed according to the action taken. Also, since the prior performance of the entity with respect to his or her change trend will bear on the validity of the difference findings, a trend factor from the trend factor table 53 is applied. If the entity's prior performance is highly variable, the validity of the difference findings is reduced. Conversely, if the prior behavior trend is found to be very stable, significant differences will be amplified.

The results of the behavior change analysis performed by the individual summary processor 51 are presented to the analyst. Where the difference results indicate a high degree of correlation with the risk criteria identified in the first review pass, the analyst can assume a causal relationship with the action and that the areas that have changed are suggestive of aberrant

behavior prior to the action. The analyst can then use the behavior changes as a means of reviewing highly specific areas of the entity's prior performance, as well as modify the weight and threshold values found in the 46 weight table and 48 threshold table to better refine the detection process for subsequent reviews. The technique embodied in the invention provides several unique and previously unavailable capabilities to the process of fraud detection.

As shown in Figs. 4, 4A, 5 and 6, details of the block flow diagrams showing the steps for generating the statistical measures for Value (v), Normalcy (n) and Change (d) are shown. As the means for generating these statistical measures are well-known in the art, they will not be described in detail herein. It should be noted, however, that although the means for generating these statistical measures are well-known, the application of these statistical measures to transactional data in order to analyze such data for purposes of fraud detection is considered a novel feature of the invention.

As shown in Fig. 7, which details a block flow diagram showing the steps for performing the pre-processor method, the initial stages of the process serve to filter and highlight suspect individuals, as described previously. Once suspect individuals are detected, prioritized and highlighted by the system an analyst will review the results and actions generated by the risklist display 40 and risk action generator 42, respectively, to determine what action to take. This automated filtering and prioritization process is critical because reviewing all of the behavior of all entities in sufficient detail to precisely identify fraud is impractical. The amount of

data normally collected in the transaction-billing process is far to great to permit manual, or human review. And the potential variations for fraud in such industries as health care are far too many, and much too flexible, to build a 5 completely automated pattern detection system that would be reliable for any protracted period. Consequently, the present invention reduces the focus of the human analyst to the most likely payoff.

10 The combined statistical approach of the pre-processor apparatus 110 highlights the most critical areas of behavior in performing the filtering, that is: value, normalcy, and change. These measures, when applied to carefully selected behavior criteria (i.e., selected for 15 their potential to exhibit profit-taking) produce a high degree of validity in showing the analyst where to look. The output of the first pass review is normally sufficient for the analyst to decide to take action.

20 This element of the present invention is one of the unique components of the process - the concept of introducing process control theory to the detection of fraud in large transaction data histories. There has not, heretofore, been available a methodology that was 25 implementable in an automated process that could record behavior changes in specific entities and determine the savings resulting from the counter-fraud program. As a result of the first pass review, organizations currently suffering from high losses due to fraud have a means of not 30 only identifying candidate targets, but they can modify the behavior of those targets in order to obtain savings and further focus on specific fraudulent behavior. The results of the process can offer the analyst with the option of

declaring savings or using the highlighted areas of change as a means of initiating deeper investigation into very specific areas of the entity's performance.

5 While the initial use of the process is focused on fraud in the health care industry, it is apparent that the technique described here can also be used to detect fraud and identify resulting savings in a variety of environments (e.g., banking, mortgage, worker's compensation, etc.).

10

15

20

25

30

I CLAIM:

1. A fraud detection system for targeting a potentially fraudulent entity within a particular industry,
5 comprising:

statistical analysis engine for generating statistical values for said entity according to at least one of a plurality of analysis criteria;

10 entity-criterion score generator for weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

15 entity score generator for weighting and summing each said entity-criterion scores to form a first entity score;

20 computer memory for storing said statistical values, said entity-criterion scores, and said first entity score; and

25 risk analysis processor for comparing said first entity score to a predetermined threshold, whereby the result of said comparison indicates whether said entity is engaging in fraudulent activity.

30 2. The system according to claim 1, further comprising controller means for coordinating, controlling, monitoring and executing commands necessary for operating said fraud detection system.

3. The system according to claim 1, further comprising a database for storing said first entity score.

4. The system according to claim 1, wherein said 5 statistical values include a value statistic (v), an normalcy statistic (n), and a change statistic (d) for each of said criteria.

5. The system according to claim 1, further 10 comprising client data converter for converting client records containing data elements useful in measuring the performance of said entity to a format compatible with said fraud detection system.

15 6. The system according to claim 1, further comprising a three dimensional array for organizing and storing said entity and said criteria according to predefined time units, said array comprising an x, y and z axis, said y-axis containing said entity, said x-axis 20 containing said criteria, and said z-axis containing said time units.

7. The system according to claim 6, wherein said 25 array creates entity-criterion buckets wherein said statistical values for each criterion can be generated for said entity within said time period.

8. The system according to claim 1, further 30 comprising a risklist display for displaying in graphical form whether said entity is engaging in fraudulent behavior.

9. The system according to claim 1, further comprising:

5 risk action generator for displaying possible actions to initiate against said entity, and selecting one of said actions to modify said entity's behavior;

10 risk behavior monitor for storing a record of said selected action taken against said entity; and

summary and savings generator for identifying the amount of money gained or lost by said action.

15 10. The system according to claim 9, wherein said summary and savings means include:

a second entity score presenting said entity's behavior after said action has occurred;

20 individual summary processor for comparing said first entity score with said second entity score, measuring the degree of change therefrom, weighting the degree of change using a severity factor and a trend factor, and displaying a result indicating whether said action decreased the fraudulent activity of said entity;

25 group summary processor for generating group statistics for a plurality of entities.

30 11. The system according to claim 10, wherein said individual summary processor includes adjuster for adjusting said weighting values for said criteria scoring means and said entity scoring means, and said threshold

values for said risk analysis means, in accordance with said group statistics from said summary and savings means.

12. A fraud detection system for targeting
5 potentially fraudulent entities within a particular
industry, comprising:

10 client data source for inputting client records
containing data elements useful in measuring the
performance of an entity by evaluating a plurality of
criteria over a set of time period;

15 there dimensional array for organizing and storing
said entity, said criteria, and said time period,
comprising an x, y and z axis, said y-axis containing said
entity, said x-axis containing said criteria, and said
z-axis containing at least one of said time units;

20 statistical analysis engine for generating a value
statistic (v), a normalcy statistic (n), and a change
statistic (d) for each of said criteria;

25 entity-criterion score generator for weighting and
summing each of said statistical values thereby forming an
entity-criterion score for each of said criteria;

entity score generator for weighting and summing
each of said entity-criterion scores to form an entity
score for said entity;

30

risk analysis processor for comparing said entity
score with pre-defined threshold values; and

risklist display for displaying in a graphical manner whether said entity is engaging in fraudulent behavior.

5 13. A computer program for targeting potentially fraudulent entities within a particular industry, said computer program comprising the steps of:

10 inputting client records containing data elements useful in measuring the performance of an entity by evaluating a plurality of criteria over a set time period;

15 organizing and storing said entity, said criteria, and said time period, into a three dimensional array comprising an x, y and z axis, said y-axis containing said entity, said x-axis containing said criteria, and said z-axis containing at least one of said time units;

20 generating a value statistic (v), a normalcy statistic (n), and a change statistic (d) for each of said criteria;

25 weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said criteria;

 weighting and summing each of said entity-criterion scores to form an entity score for said entity;

30 comparing said entity score with pre-defined threshold values; and

displaying graphically whether said entity is engaging in fraudulent behavior.

14. The computer program according to claim 13, 5 said computer program further comprising the step of encoding said computer program on a computer-readable medium.

15. A method of targeting potentially fraudulent 10 entities within a particular industry, comprising the steps of:

generating statistical values for entity according to a plurality of analysis criteria;

15 weighting and summing each of said statistical values thereby forming an entity-criterion score for each of said analysis criteria;

20 weighing and summing each of said entity-criterion scores to form an entity score; and

comparing said entity score to a predetermined threshold.

25 16. The computer program according to claim 15, said computer program further comprising the step of encoding said computer program on a computer-readable medium.

30 17. A fraud pre-processor system for filtering and highlighting those individuals most likely to be engaging in fraudulent behavior, comprising:

a controller;

a computer program executed by said controller comprising:

5

client data converter for converting client data to format compatible with said fraud pre-processor system;

10

three dimensional array for storing a plurality of entities, analysis criteria, and time units such that for each of said entities statistical values for each of said criteria within each of said time units can be generated, weighted and summed to form an entity-criterion score;

15

entity-criterion score generator for generating said entity-criterion score;

20

entity score generator for weighting and summing each entity-criterion score for each entity to form an entity score;

25

risk analysis processor for comparing said entity score for each of said entities with a threshold value;

30

risklist display for displaying the results from said comparison in a manner identifying those entities most likely engaging in fraudulent behavior; and

35

a database for permanently storing, and a computer memory for temporarily storing, said analysis criteria, said statistical values, said entity-criterion score, said entity-score, said threshold values, and said results.

18. A fraud detection system, comprising statistical analysis engine for generating statistics for a plurality of entities according to a plurality of analysis criteria and time periods; and

5

pre-processor for evaluating said statistics to determine if said entities are engaging in fraudulent behavior.

10

19. The system according to claim 1, wherein said statistical analysis engine further comprises:

value processor for generating said value statistic (v) which determines the damage potential for said entity;

15

normalcy processor for generating said normalcy statistic (n) for measuring normal behavior for said entities; and change processor for generating said change statistic (d).

20

20. A fraud detection system, comprising:

risk analysis processor for comparing an entity score for an entity to a predetermined threshold;

25

risk action generator for displaying and selecting action to be taken said entity;

30

risk action method for storing a record of said actions; summary and savings generator for identifying the amount of money gained or lost by said action; and

adjustor for adjusting weighting values used to generate said entity score, and said threshold values for said risk analysis means, in accordance with said statistics from said summary and savings generator, in order to better estimate if said entity is engaging in fraudulent behavior.

10

15

20

25

30

1/9

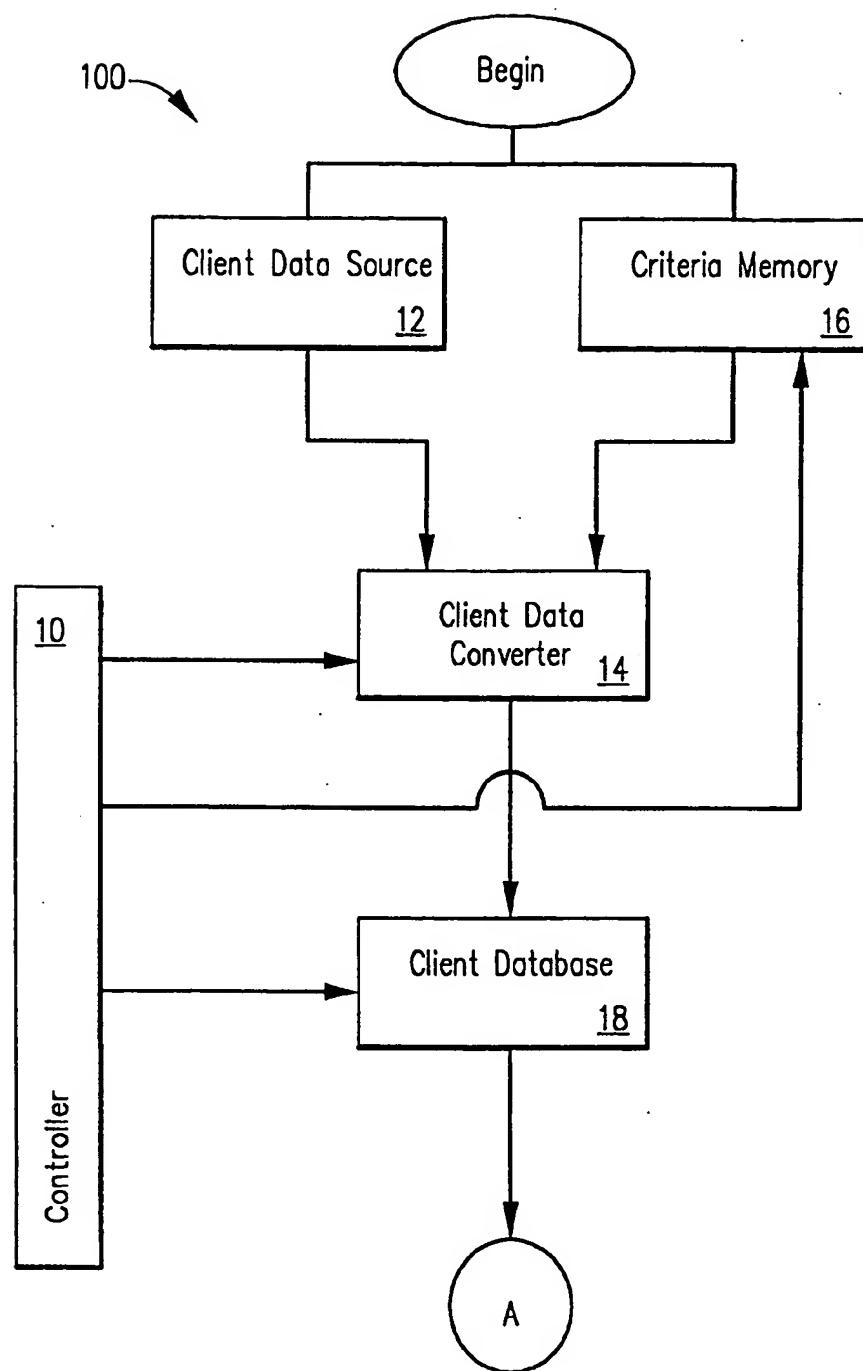


FIG. 1

2/9

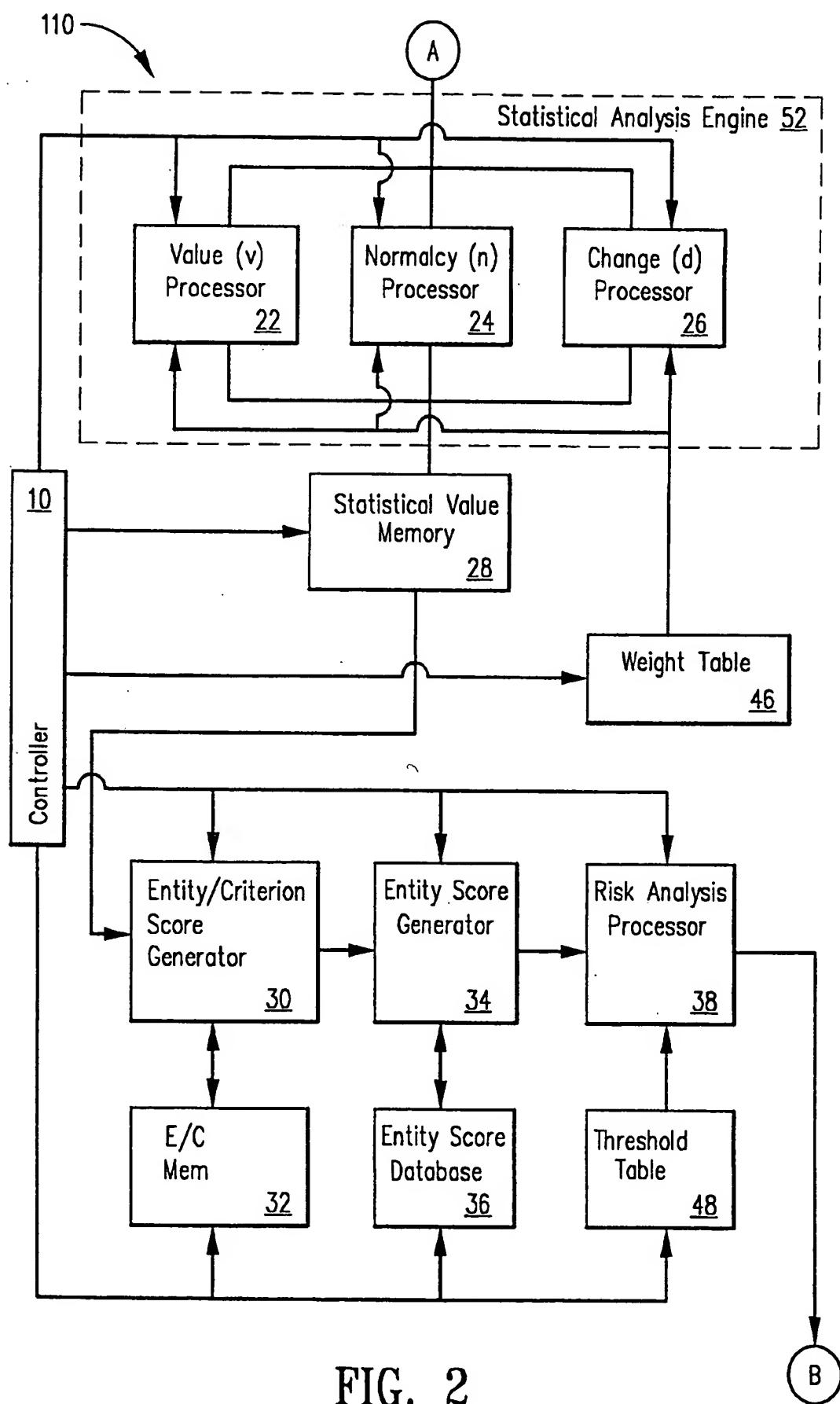


FIG. 2

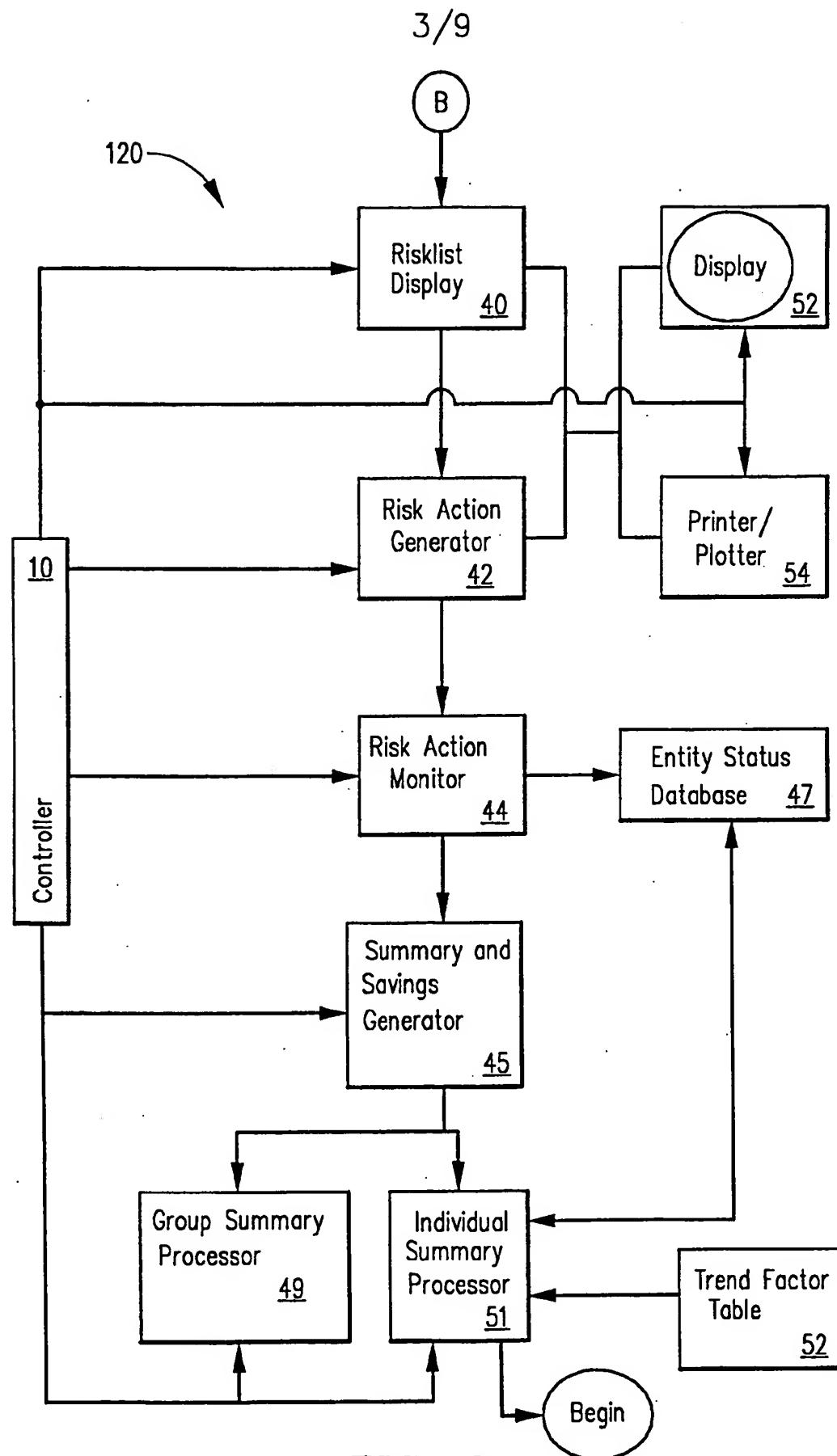
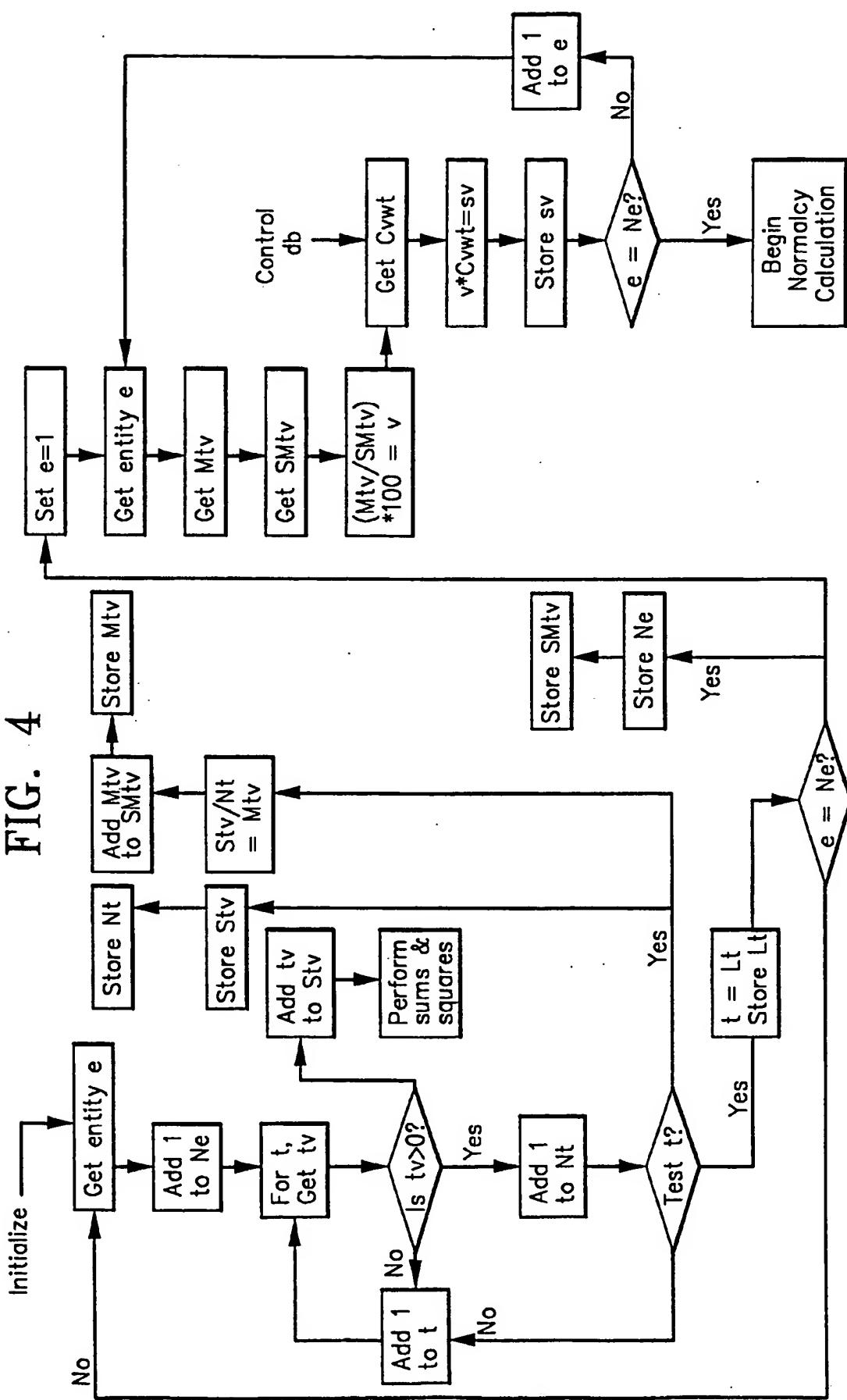


FIG. 3

4 / 9



5/9

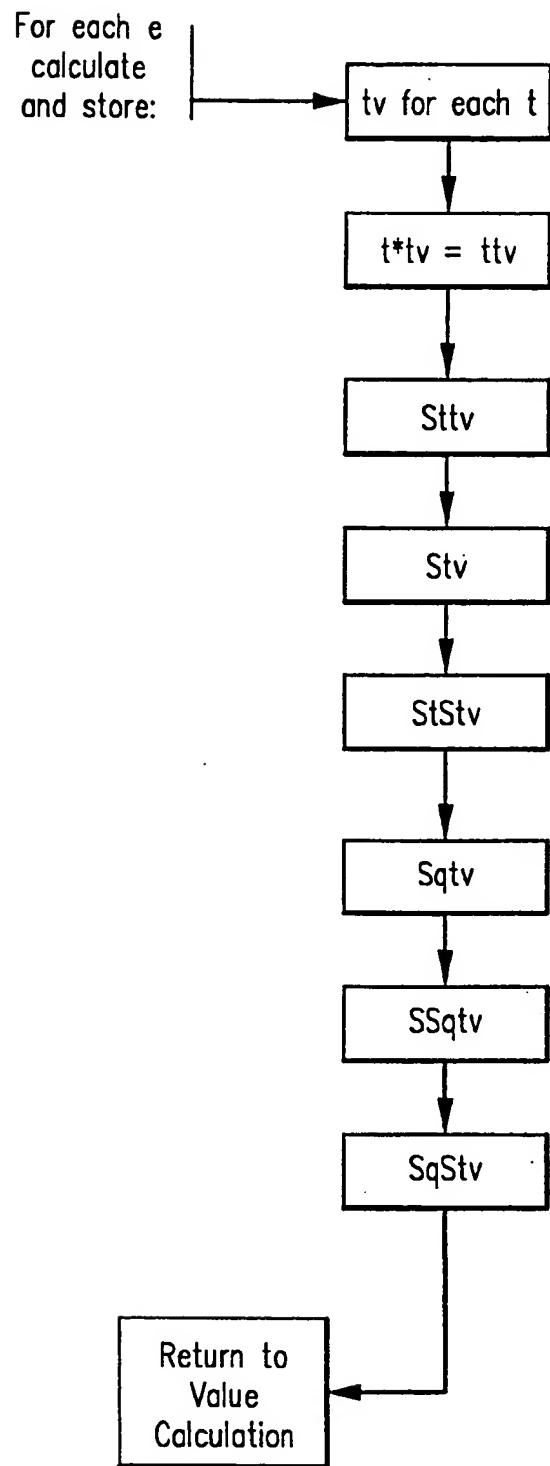
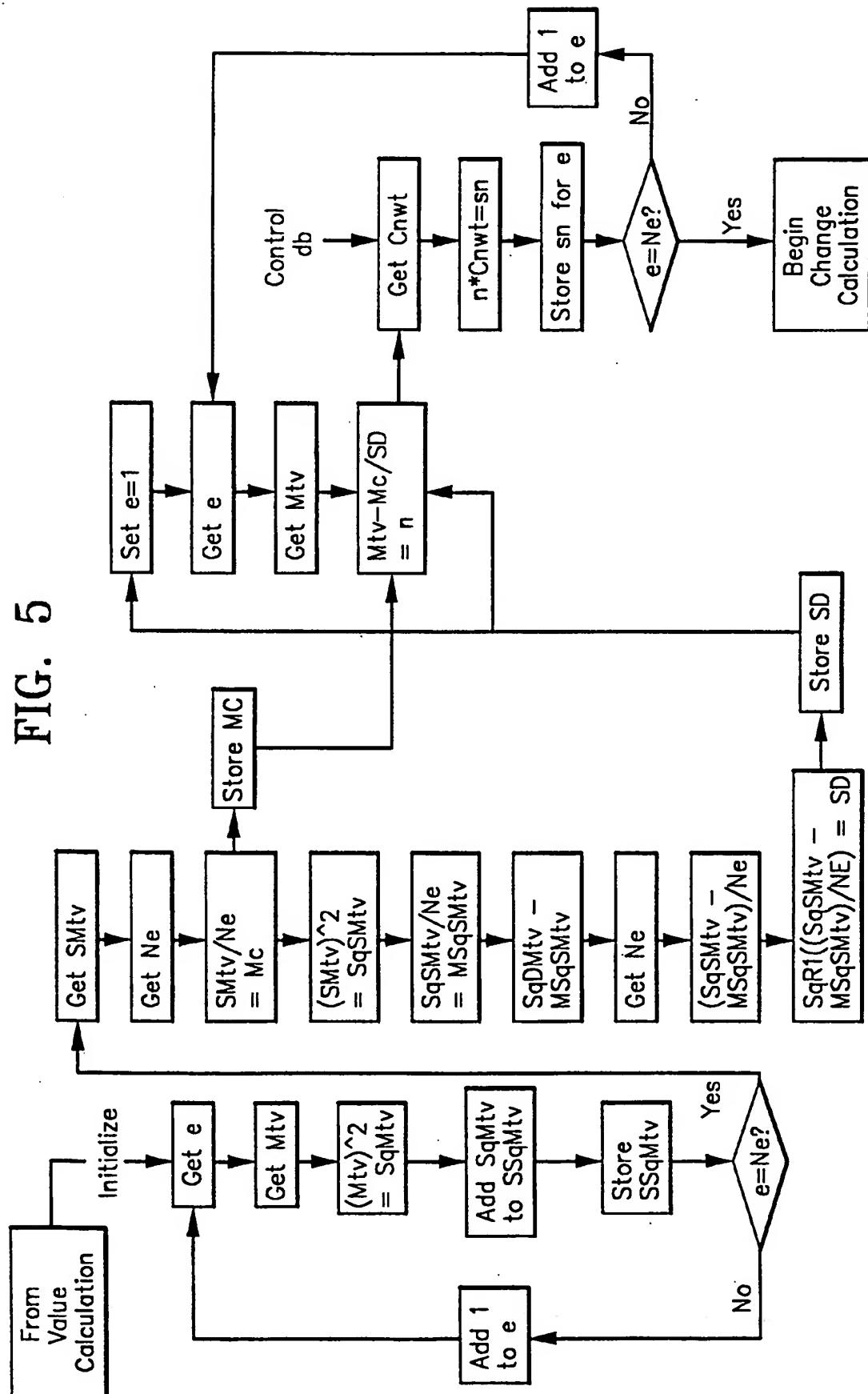


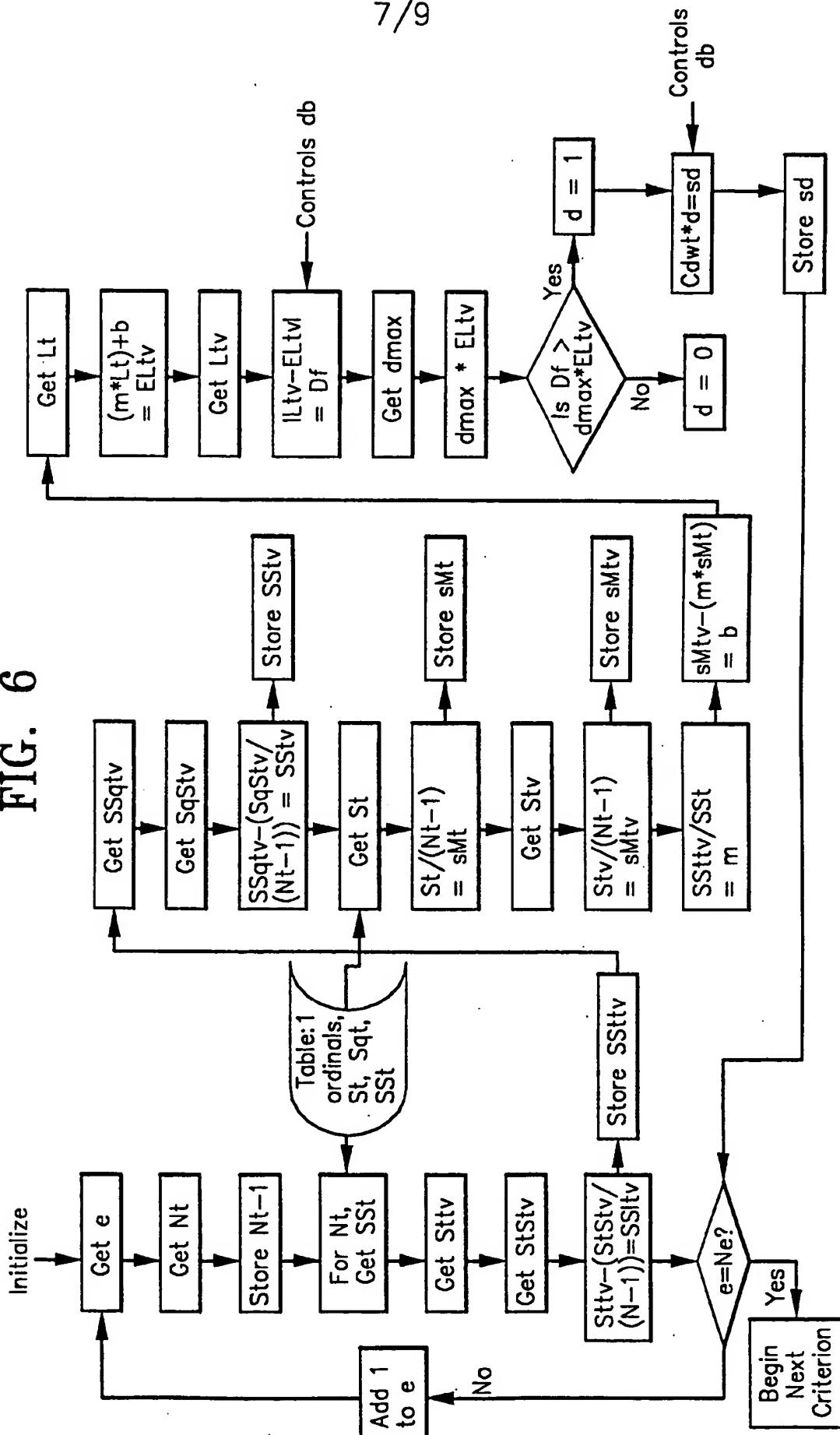
FIG. 4A

6/9

FIG. 5



7/9



8/9

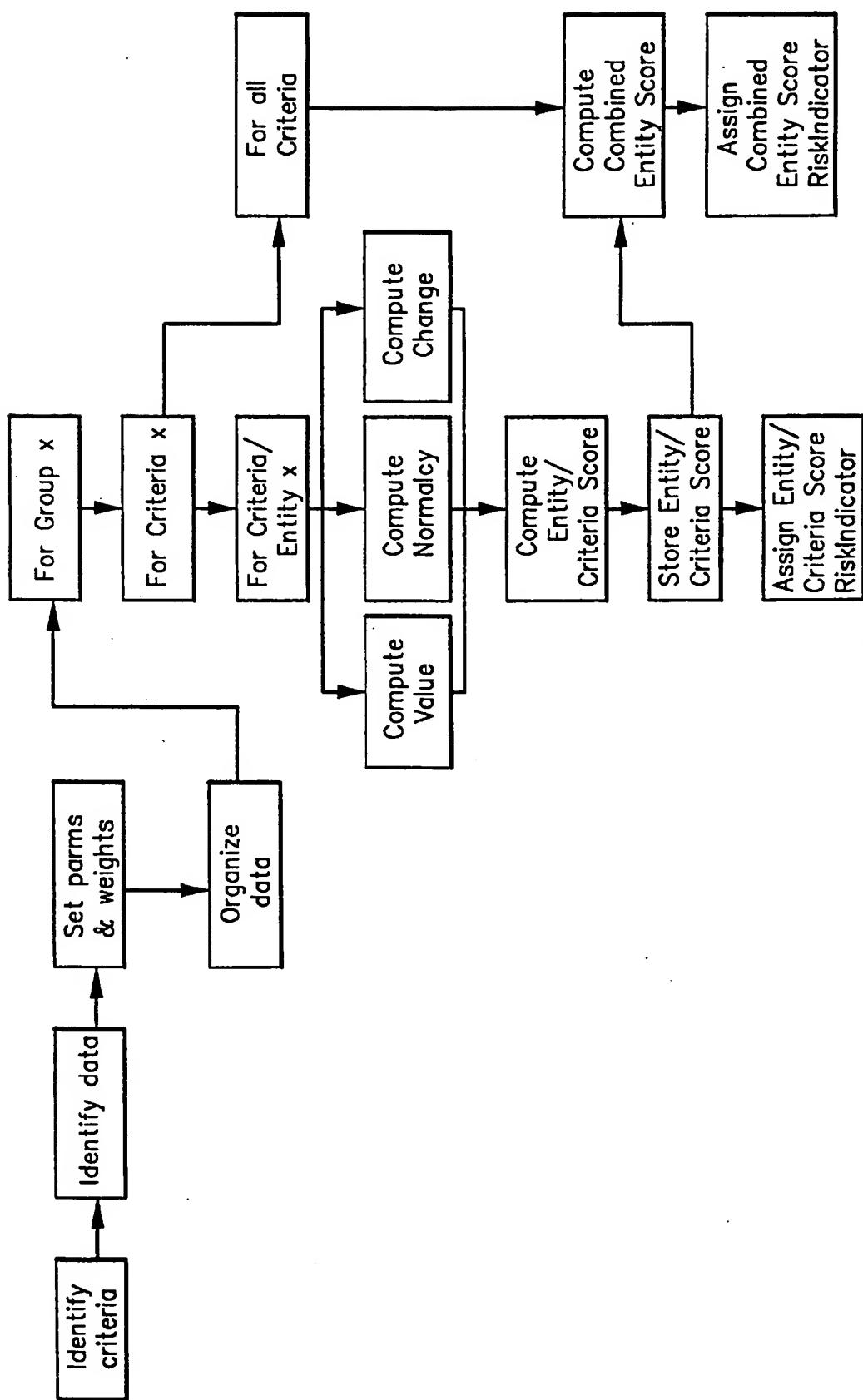


FIG. 7

9/9

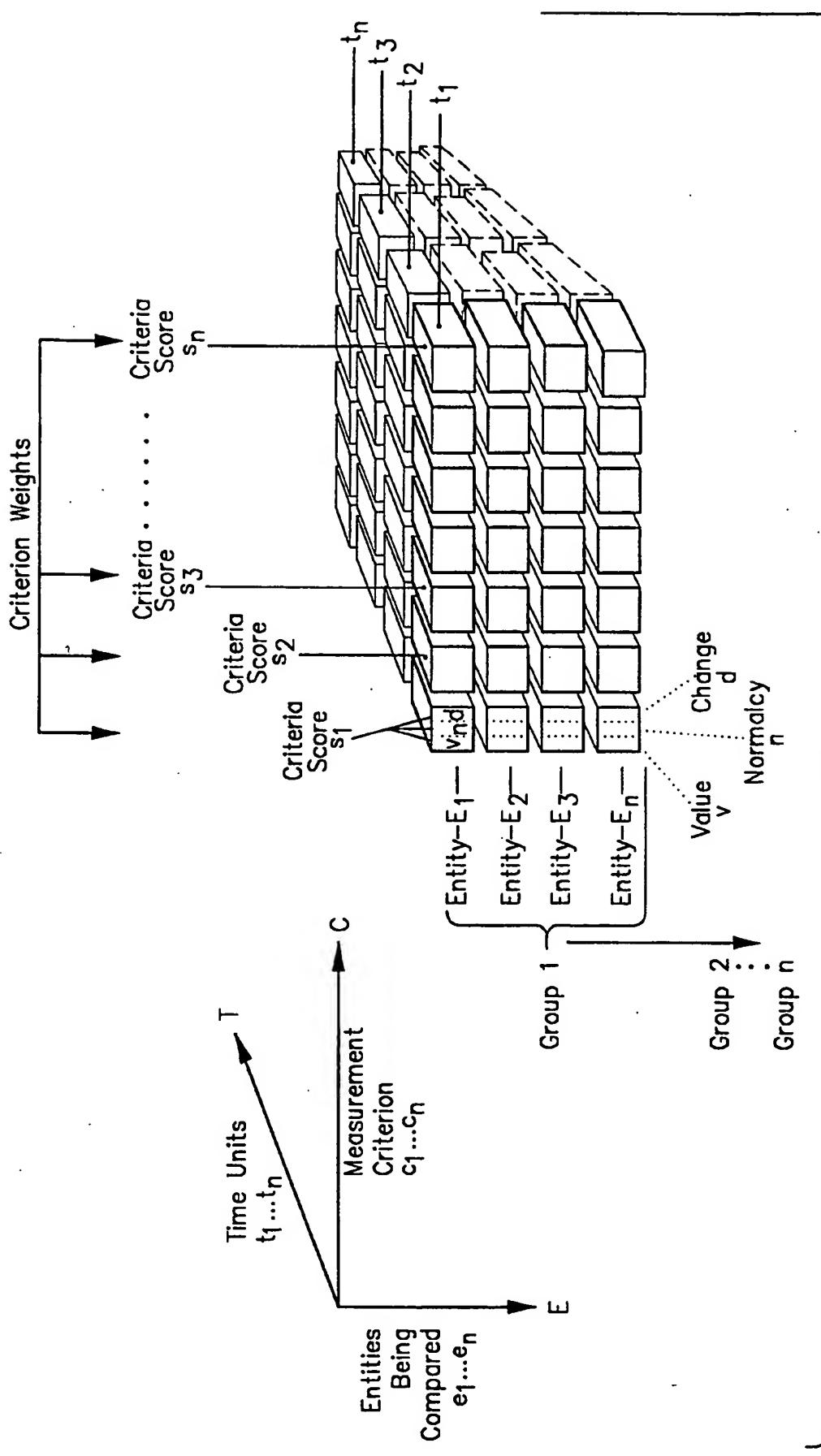


FIG. 8